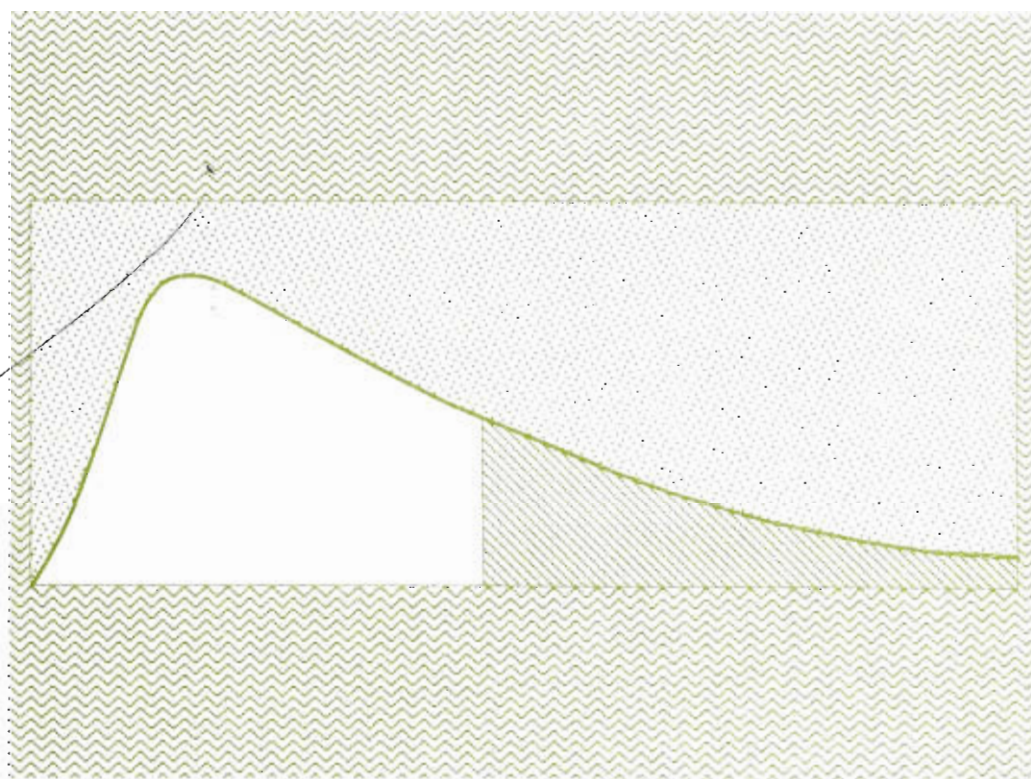


864

Guía SAI de **probabilidad y estadística**

Tronco general

Jaime Grabinsky



UAM
QA276.12
G7.4
2001

UNIVERSIDAD
AUTÓNOMA
METROPOLITANA
Casa abierta al tiempo **Azcapotzalco**

Guía SAI de
**probabilidad
y estadística**
Tronco general

Guía SAI de
**probabilidad
y estadística**
Tronco general

Jaime Grabinsky



2893128

UAM-AZCAPOTZALCO

RECTOR

Mtro. Víctor Manuel Sosa Godínez

SECRETARIO

Mtro. Cristian Eduardo Leriche Guzmán

COORDINADOR GENERAL DE DESARROLLO ACADÉMICO

Mtro. Luis Soto Walls

COORDINADORA DE EXTENSIÓN UNIVERSITARIA

DCG Ma. Teresa Olalde Ramos

JEFA DE LA SECCIÓN DE PRODUCCIÓN Y DISTRIBUCIÓN EDITORIALES

Lic. Silvia Graciela Lona Perales

ISBN: 970-654-633-2

© UAM-Azcapotzalco

Jaime Grabinsky

Universidad Autónoma Metropolitana
Unidad Azcapotzalco
Av. San Pablo 180
Col. Reynosa Tamaulipas
Delegación Azcapotzalco
C.P. 02200
México, D.F.

Sección de producción
y distribución editoriales
Tel. 5318-9222/9223
Fax. 5318-9222

2a. edición, 2001
1a. reimpresión, 2002

Impreso en México.

I
T E M A S

1. INTRODUCCIÓN
2. ESTADÍSTICA DESCRIPTIVA
3. PROBABILIDAD
4. MODELOS PROBABILÍSTICOS
5. INFERENCIA ESTADÍSTICA

II

1. INTRODUCCIÓN (1 sesión)
 - 1.1 Necesidad de la estadística.
 - 1.2 Introducción a la aleatoriedad y al determinismo.
 - 1.3 Aplicaciones en Ingeniería.

2. ESTADÍSTICA DESCRIPTIVA (3 sesiones)
 - 2.1 Definición. (1)
 - 2.2 Organización de Datos.
 - 2.2.1 Conceptos fundamentales.
 - 2.2.1.1 Conceptos de población, unidades elementales, observación y muestra.
 - 2.2.1.2 Parámetros y estadísticas.
 - 2.2.1.3 Tipos de información.
 - 2.2.2 Distribuciones de Frecuencias.
 - 2.2.2.1 Frecuencia absoluta, relativa y acumulada.
 - 2.2.2.2 Densidad de frecuencias.
 - 2.2.3 Representación Tabular y Gráfica.
 - 2.2.3.1 Histogramas y polígonos de frecuencia.
 - 2.2.4 Modelos de Población.
 - 2.3 Medidas de Tendencia Central. (1)
 - 2.3.1 Media.
 - 2.3.1.1 Aritmética simple y ponderada.
 - 2.3.1.2 Para datos simples y datos agrupados.
 - 2.3.1.3 Proporciones, cuartiles y percentiles.
 - * 2.3.1.4 Otros. (Geométrica, Armónica, Cuadrática).
 - 2.3.2 Mediana.
 - 2.3.2.1 Para datos simples y datos agrupados.
 - 2.3.3 Moda.
 - 2.3.3.1 Datos simples y datos agrupados.
 - 2.4 Medidas de Dispersión, Asimetría y Aplanamiento. (1)
 - 2.4.1 Dispersión.
 - 2.4.1.1 Rango.

III

- 2.4.1.2 Desviación media.
- 2.4.1.3 Desviación standard y variancia.
- 2.4.1.4 Datos simples y agrupados.
- 2.4.1.5 Coeficiente de Variación.
- * 2.4.2 Medidas de Asimetría.
- * 2.4.3 Medidas de Aplanamiento.

3. PROBABILIDAD

(9 sesiones)

- 3.1 Introduucción.
 - 3.1.1 Medición del riesgo.
 - 3.1.2 Importancia.
 - 3.1.2.1 Intrínseca.
 - 3.1.2.2 Para la inferencia estadística.
- 3.2 Definición.
 - 3.2.1 Clásica.
 - 3.2.1.1 Axiomas de la probabilidad.
 - 3.2.2 Empírica (frecuencia relativa).
 - 3.2.3 Subjetiva.
 - 3.2.3.1 Importancia práctica.
- 3.3 Repaso de conjuntos.
 - 3.3.1 Definiciones.
 - 3.3.2 Notación.
 - 3.3.3 Algebra de conjuntos.
 - 3.3.4 Diagramas de Venn.
- 3.4 Experimentos, Espacios muestrales y Eventos.
 - 3.4.1 Definiciones.
 - 3.4.2 Tipos de eventos.
 - 3.4.2.1 Mutuamente exclusivos y translapados.
 - 3.4.2.2 Complementarios.
 - 3.4.2.3 Independientes.
- 3.5 Técnicas de Conteo y Cálculo de Probabilidades.
 - 3.5.1 Principio de Adición y Multiplicación.
 - 3.5.2 Combinaciones.
 - 3.5.3 Permutaciones.

IV

- 3.5.4 Ley Aditiva de la Probabilidad.
- 3.5.5 Cálculo de probabilidades.
- 3.6 Probabilidad Condicional e Independencia.
 - 3.6.1 Ley Multiplicativa de la Probabilidad.
 - 3.6.2 Arboles de probabilidad.
 - 3.6.3 Probabilidad condicional.
 - 3.6.4 Independencia.
 - 3.6.5 Teorema de la Eliminación (Probabilidad total).
 - 3.6.6 Teorema de Bayes.
- 3.7 Variables Aleatorias.
 - 3.7.1 Definición.
 - 3.7.2 Distribuciones de probabilidad.
 - 3.7.2.1 Discretas y continuas.
 - 3.7.2.2 Totales y acumuladas.
 - 3.7.3 Esperanza Matemática.
 - 3.7.3.1 Cálculo.
 - 3.7.3.2 Propiedades.
 - 3.7.4 Variancia.
 - 3.7.4.1 Cálculo.
 - 3.7.4.2 Propiedades.
- * 3.8 Probabilidad Conjunta y Marginal.
 - 3.8.1 Variables bidimensionales.
 - 3.8.2 Probabilidad conjunta.
 - 3.8.3 Probabilidad marginal.
 - 3.8.4 Covariancia.

4. MODELOS PROBABILÍSTICOS

(5 sesiones)

- 4.1 Introducción.
 - 4.1.1 Descripción.
 - 4.1.2 Modelo (Definición).
 - 4.1.3 Modelo Probabilístico.
 - 4.1.4 Tipos de modelos.
 - 4.1.4.1 Discretos.
 - 4.1.4.2 Continuos.

4.2 Modelos Discretos.

4.2.1 Rectangular.

4.2.1 Bernoulli.

4.2.2 Binomial.

4.2.3 Poisson.

* 4.2.4 Otros.

4.3 Modelos Continuos.

4.3.1 Uniforme.

4.3.2 Normal.

4.3.2.1 Conceptos básicos.

4.3.2.2 Importancia.

4.3.2.3 Normal estandarizada.

4.3.2.4 Aplicaciones.

* 4.3.3 Aproximaciones de Binomial y Poisson.

4.3.4 Exponencial.

* 4.3.5 Relación entre Exponencial y Poisson.

5. INFERENCIA ESTADÍSTICA

(12 sesiones)

5.1 Teoría del Muestreo.

(3)

5.1.1 Introducción.

5.1.1.1 Necesidad.

5.1.1.2 Tipos de muestreo.

5.1.1.3 Errores de estimación.

5.1.2 Observaciones como variables aleatorias.

5.1.3 Suma de variables aleatorias.

5.1.4 Teorema del Límite Central.

5.1.5 Distribuciones muestrales.

5.1.5.1 Media.

5.1.5.2 Proporciones.

5.1.5.3 Variancia.

5.1.5.4 Diferencias.

* 5.1.6 Ley de los Grandes Números.

* 5.1.7 Desigualdad de Chebyshev.

5.2 Estimación.

(3)

5.2.1 Concepto de Estimación.

VI

- 5.2.1.1 Inferencia estadística.
- 5.2.1.2 Estimación puntual y por intervalo.
- 5.2.2 Estimación Puntual.
 - 5.2.2.1 Características de los estimadores.
 - * 5.2.2.2 Métodos de obtención de estimadores.
 - 5.2.2.3 Estimadores de máxima verosimilitud.
- 5.2.3 Estimación por Intervalo.
 - 5.2.3.1 Intervalos de confianza para media, variancia y diferencias.
 - 5.2.3.2 Errores de estimación.
 - 5.2.3.3 Tamaño mínimo de la muestra.
- 5.3 Pruebas de Hipótesis. (6)
 - 5.3.1 Conceptos.
 - 5.3.1.1 Formulación de hipótesis.
 - 5.3.1.2 Tipos de error y probabilidades asociados.
 - 5.3.1.3 Reglas de decisión.
 - 5.3.2 Selección del tipo de prueba estadística.
 - 5.3.2.1 Estadística de prueba.
 - 5.3.2.2 Modelos probabilísticos.
 - 5.3.2.3 Alternativas de una y dos colas.
 - 5.3.3 Pruebas de Hipótesis.
 - 5.3.3.1 Para medias con μ conocida.
 - 5.3.3.2 Para medias con σ desconocida.
 - 5.3.3.3 Para diferencia de medias.
 - 5.3.3.4 Para proporciones.
 - 5.3.4 Relaciones entre μ , σ y el tamaño de la muestra.

T O T A L 30 sesiones

NOTA: Los subtemas marcados con asterisco (*), se darán si el tiempo para esos temas lo permite.

ÍNDICE

INTRODUCCIÓN	1
Unidad I PROBLEMAS DE DECISIÓN ESTADÍSTICA.....	3
Unidad II DISTRIBUCIONES DE FRECUENCIAS.....	7
Unidad III MEDIDAS DE TENDENCIA CENTRAL.....	9
Unidad IV MEDIDAS DE VARIACIÓN Y ASIMETRÍA.....	11
Unidades V y VI PROBABILIDAD Y VARIABLES ALEATORIAS	13
Unidad V PROBABILIDAD.....	15
Unidad VI DISTRIBUCIONES DE PROBABILIDAD.....	19
Unidad VII MODELOS PROBABILÍSTICOS ESPECIALES.....	27
Unidades VIII, IX, X INFERENCIA ESTADÍSTICA.....	31
Unidad VIII TEORÍA DEL MUESTREO.....	33
Unidad IX ESTIMACIÓN.....	37
Unidad X DOCIMASIA ESTADÍSTICA.....	41
PRIMERA UNIDAD DE REVISIÓN.....	45
SEGUNDA UNIDAD DE REVISIÓN.....	49
Y.L. Chou. TABLAS.....	51

PROBABILIDAD Y ESTADÍSTICA

UNIDADES I, II, III y IV

INTRODUCCIÓN Y ESTADÍSTICA DESCRIPTIVA

Estas primeras unidades básicamente 'describen' los resultados de experimentos. Si se revisa el programa de estudios de la materia, se observará que debe llevar un máximo de 4 sesiones. Recomendamos que se avance muy rápidamente en ellas y de ser posible, se lean el Huff y/o el Moroney (por lo menos ver los dibujos), que ilustran las complicaciones que encierran estos conceptos, ya en la práctica. También recomendamos la lectura del anexo "The role of Probability Theory in Engineering" así como los resúmenes de artículos extraídos del Engineering Index. Creemos que con esto queda más que demostrado el uso práctico de estos conocimientos.

Los objetivos de estas unidades son conocer las fórmulas correspondientes a los 'promedios' y a las medidas de dispersión, y su interpretación gráfica. Calcular propiedades de muestras cuando a las medidas se les ha hecho "algún cambio", desarrollando cierta habilidad algebraica para ello; poder trasladar estos conceptos teóricos a situaciones extraídas de la vida real, sabiendo escoger los parámetros de genuino interés.

NOTA IMPORTANTE

La primera unidad consiste en los primeros cuatro capítulos de la 1a. edición del Chou. Lo fundamental para presentar el examen de la 1a. Unidad es estudiar los problemas de Estadística Descriptiva del problemario.

PROBABILIDAD Y ESTADÍSTICA

UNIDAD I

PROBLEMAS DE DECISIÓN ESTADÍSTICA

Este es un capítulo introductorio donde se busca dar sentido y razón de ser a las técnicas más precisas que se estudiarán adelante. La investigación estadística responde a la necesidad de resolver problemas concretos, en ocasiones urgentes y de vital importancia.

Por ello, antes de embarcarse en un estudio estadístico es necesario plantearse con precisión las características y fines del problema que se busca enfrentar y resolver. La elección de las técnicas, de las medidas cuantitativas asociadas al problema, dependen de la claridad en la concepción, de la calidad del esfuerzo en este primer paso. Siempre existe el peligro de dejarse confundir con la multiplicidad de recursos que se ofrecen pero sólo una voluntad bien informada, consciente, evita ser atrapada.

Estos comentarios son válidos también para formarse un criterio inteligente y evaluar la multiplicidad de trabajos estadísticos que demanden nuestra atención.

Procedimiento

1. Lea las siguientes secciones del libro, 1-1 a la 1-11, págs. 1-21.
2. Resuelva los siguientes problemas pág. 21: 1-5, 1-13, 1-14, 1-16, 1-17, 1-18, 1-19, 1-20, 1-21.
3. Lea la Introducción del libro How to Lie with Statistics Huff.

Autoevaluación

1. Haga un resumen exhaustivo de los temas tratados en el capítulo, con sus propias palabras y sin ver el libro. Al terminar compare y complete o corrija.
2. Describa 5 situaciones muy cercanas a la vida diaria real, describa problemas y formule un plan de decisión basado en estudios estadísticos.
3. Las leyes de cada materia científica son ejemplos de decisiones estadísticas muy exitosas. Ilustre el método estadístico con algunas leyes importantes de Ingeniería, Física o Química. Comente sobre la muy íntima interrelación entre la Teoría pertinente y Estadística.
4. Dé el máximo de propiedades que pueda, referentes a la población de estudiantes de la UAM y a la población de personal académico. Suponga que se está estudiando la posibilidad de mejorar la educación que reciben los estudiantes y de reducir al mínimo la deserción y los fracasos escolares. ¿Cuáles de las propiedades anteriores serían útiles para este estudio? ¿Qué parámetros de ellas?
5. Suponga que usted tiene que decidir entre (a) dedicar más recursos a la extensión del Seguro Social en el campo, con lo que sabe por experiencia previa que disminuirá la mortalidad (de 300 a 200 por año aproximadamente en una comunidad de 10,000 habitantes), (b) dedicarlo a mejorar la biblioteca del Centro Médico. ¿Qué estudios mínimos elaboraría para evaluar el impacto de las acciones y escoger entre ellas?
6. Un ingeniero cree intuitivamente que la válvula que va a instalar en la Refinería de Azcapotzalco, va a funcionar y no va a explotar. ¿Qué tipo de estudio para reforzar su creencia considera necesario?

7. Frecuentemente hay más teorías que verdades referentes a muchos hechos objetivos. La estadística la hace frecuentemente de juez entre estas teorías en competencia. Pero en muchas ocasiones el que elabora estadísticas sólo busca tomar partido y cobijarse bajo la pretendida "imparcialidad" deformando éstas para bien de sus fines. Por ejemplo, un defensor de un producto puede mencionar una muestra en que su producto mostró ventajas, sin relacionarlo con las características de otras muestras o de la población. Un ejemplo clásico fue el de la talidomida que tan caro le resultó al laboratorio que la elaboró. Mencione otro tipo de falacias posibles o que conozca.
8. Se sabe que gente expuesta en forma continua a un ruido que pasa determinado umbral, medido en decibeles, adquiere problemas auditivos, con frecuencia irreversibles. ¿Qué estudio haría en un lugar de trabajo para evitar este daño?
9. Se sabe o se dice que los europeos tienden a deducir (inducir) lo máximo posible partiendo de un conjunto de datos experimentales; mientras que los americanos tienden a inducir poco y a experimentar más. Comente sobre las posibles ventajas y desventajas de ambas estrategias.
10. En muchos experimentos de Percepción Extra Sensorial, los "científicos" han decidido rechazar procedimientos y análisis estadísticos. ¿Qué tipo de problemas se generan al efectuar estas investigaciones?
11. Hay muchas variables asociadas al estudiar matemáticas bien, por ejemplo, hacer ejercicios, redemostrar teoremas, hacer ejercicios teóricos, rechazar la teoría, saber la historia, rehacer las definiciones, analizar aplicaciones, hacer resúmenes, etc. Asócielas a estas variables aleatorias la distribución en su caso personal y vea la interrelación entre promedios y el aprendizaje.

12. Una máquina llenadora de latas de alimentos no siempre surte con la misma cantidad de producto a cada lata. Al industrial, al consumidor y al inspector fiscal les interesa mucho saber este comportamiento, para evitar fraudes, "regalos" al consumidor o violaciones a los reglamentos. ¿Cuáles son los parámetros estadísticos de interés?

UNIDAD II

DISTRIBUCIONES DE FRECUENCIAS

La utilidad de los medios visuales es innegable para ilustrar conceptos y reflejar hechos. Pero la variedad de estos hace necesaria su elección consciente. Varias de las representaciones más naturales de las propiedades de una muestra se estudian en este capítulo, formalizando y precisando puntos referentes a gráficas con las que tenemos contacto en múltiples circunstancias. (Mucho esfuerzo e ingenio se ha empleado en esto).

Procedimiento

1. Estudiar el texto, el capítulo 2.
2. Resuelva los siguientes problemas de su texto.
2-4, 2-13, 2-21, 2-23, 2-25 y 2-27.
3. Lea los capítulos 1 y 2 del Huff.
4. Resuelva la Autoevaluación.

Autoevaluación

1. Defina los siguientes conceptos: percentil, rango percentil, escala, calificación, histograma, tabla de frecuencias, polígono de frecuencias, distribución acumulativa.
2. ¿Cómo interpreta el área bajo un histograma? ¿Existe alguna interpretación del área bajo la curva de porcentaje acumulado?

3. ¿Cómo pasaría la información contenida en un histograma a un círculo dividido en sectores?
4. Haga un resumen exhaustivo de los temas tratados en el capítulo con sus propias palabras y sin ver el libro, al terminar compare, complete o corrija.
5. A través del libro "How to lie with Statistics" de Huff se presentan usos deformados intencionalmente de la presentación gráfica de datos estadísticos. Describa algunos de estos.
6. Defina y describa los siguientes conceptos: series temporales, multivariante, distribución de frecuencias, límites de las clases, densidad de frecuencias, límites de clases nominales y límites de clases reales, intervalo de clases de extremos abiertos, polígono de frecuencias, estadísticas analíticas, modelos de población, curva normal o campana de Gauss, frecuencias acumulativas, distribución de frecuencias relativas, distribución de frecuencias acumulativas relativas, distribución de frecuencias absolutas y distribución de frecuencias categóricas.
7. Los ingresos de 200 estudiantes universitarios que trabajan tiempo parcial están distribuidos como sigue:

Ingresos Semanales:

77.5	82.5	5
82.5	87.5	12
87.5	92.5	13
92.5	97.5	22
97.5	102.5	30
102.5	107.5	35
107.5	112.5	32
112.5	117.5	20
117.5	122.5	15
122.5	127.5	10
127.5	132.5	6
<hr/>		
TOTAL=200		

Grafique esta tabla, dibuje el histograma, el polígono y la distribución acumulativa.

UNIDAD III

MEDIDAS DE TENDENCIA CENTRAL

En esta unidad se estudian la media aritmética, la mediana y la moda. Estos números representan, de algún modo, a la muestra que se está estudiando. Por otro lado, no todas sirven necesariamente para los mismo propósitos. Es necesario diferenciarlas conceptualmente con rigor.

La confusión de ellas en muchos casos es peligrosa y puede ser deliberada.

Procedimiento

1. Leer del capítulo III, secciones 3-1 a 3-5 y 3-10 a 3-14.
2. Resolver los siguientes problemas: 3-1, 3-2, 3-3, 3-7, 3-12, 3-13, 3-15, 3-16, 3-17, 3-21 y 3-23.
3. Lea los capítulos 3, 4 y 5 del Huff.

Autoevaluación

1. Haga un resumen exhaustivo de los temas tratados en el capítulo con sus propias palabras y sin ver el libro. Al terminar compare y complete o corrija.
2. Defina y/o describa: media aritmética, desviación media, media aritmética ponderada, mediana, clase mediana, moda, clase modal.

3. Exprese lo siguiente en notación sumatoria:

$$\left(\frac{x}{y} - 1\right) + \left(\frac{x}{y} - 1\right) + \dots + \left(\frac{x}{y} - 1\right)$$

4. Calcule las diversas medidas de tendencia central y la varianza para los siguientes datos de tamaños de tronco de madera que llegan a un aserradero en un río:

<u>PUNTO MEDIO DEL INTERVALO</u>	<u>FRECUENCIA</u>
1.30	1
1.35	5
1.40	6
1.45	13
1.50	8
1.55	17
1.60	4
1.65	7
1.70	1
1.75	3
T O T A L	<u>75</u>

5. En cierto viaje un automovilista adquirió 10 galones de gasolina a 29 centavos por galón, 15 galones a 31 centavos por galón, 20 galones a 33 centavos por galón ¿cuál es el precio medio por galón de gasolina pagado por el automovilista?

PROBABILIDAD Y ESTADÍSTICA

UNIDAD IV

MEDIDAS DE VARIACIÓN Y ASIMETRÍA

La dispersión o la variabilidad en los valores de una muestra está siempre presente. Su importancia es fundamental.

Por ejemplo un tornillo micrométrico se produce y mide .01 milímetro y se nos dice que la "variación" es grande. Los que usen este tornillo no quieren tal "variación" ya que no podrían usarlo en la producción de microscopios o igualmente, no es lo mismo producir camisas para cubrir las necesidades de una población homogénea que para una que no lo es. Para comprobar la teoría de la relatividad era crucial que la "variación" debida al diseño experimental al medir la velocidad de la luz al pasar por las cercanías del sol fuera muy pequeña. De otro modo el experimento no dilucidaba la exactitud de la teoría.

Por otro lado una población en que la mayoría tiene un nivel de vida adecuado o en que la mayoría no lo tiene, difieren fuertemente en muchas características sociales. Precisar matemáticamente el concepto de asimetría que nos permita diferenciar estas dos situaciones es parte de esta unidad. Una revolución social tiene como meta cambiar la asimetría de la distribución de ingreso (de positiva a negativa, estadísticamente) de negativa a positiva socialmente.

Procedimiento

1. Estudiar del capítulo 4, secciones: 4-1, 4-2, 4-3, 4-4 y 4-6 y resolver 4; 4-5, 4-9, 4-12, 4-13 y 4-17.
2. Leer los capítulos 6, 7 y 8 del Huff.

Autoevaluación

1. Haga un resumen exhaustivo de los temas tratados en el capítulo con sus propias palabras y sin ver el libro. Al terminar compare y complete o corrija.
2. Defina y/o describa: dispersión o variación, asimetría, leptomeso y platicúrtica, fluctuación intercuartil, fluctuación semiintercuartil, desviación media, variancia muestral, medida standard, medida de variación relativa, coeficiente de variación, coeficiente de desviación cuartil, medida pearsoniana de asimetría.
3. Para los siguientes datos calcule una tabla de frecuencias con intervalos de clase de 0.02 empezando con 6.60-6.61, 6.62-6.63, etc.; dibuje un histograma. Calcule la media y la desviación standard para los datos y para los datos agrupados.

Diámetros de Cabezas de Remaches en Centésimas de Pulgadas.

6.72	6.77	6.82	6.70	6.78	6.70	6.62
6.75	6.66	6.66	6.64	6.76	6.73	6.80
6.72	6.76	6.76	6.68	6.66	6.62	6.72
6.76	6.70	6.78	6.76	6.67	6.70	6.72
6.74	6.81	6.79	6.78	6.66	6.76	6.72
6.74	6.70	6.78	6.76	6.70	6.76	6.72
6.67	6.62	6.68	6.74	6.74	6.81	6.76
6.68	6.72	6.74	6.64	6.79	6.72	6.66
6.80	6.74	6.73	6.81	6.77	6.60	6.82
6.68	6.78	6.76	6.74	6.70	6.64	6.72
6.72	6.71	6.64	6.70	6.70	6.75	6.78
6.67	6.72	6.76	6.64	6.69	6.73	
6.62	6.66	6.84	6.73	6.66	6.66	
	6.72	6.80	6.72	6.76	6.72	

4. ¿Qué sucede a la media y a la desviación standard si a estas medidas se les agrega una constante? Si se les divide entre C?
5. Pruebe la identidad algebraica $\sum_{i=1}^N (x_i - \bar{x}) = \sum_{i=1}^N x_i - N\bar{x}$.

Exprésela en forma matricial.

UNIDADES V Y VI

PROBABILIDAD Y VARIABLES ALEATORIAS

Estas dos unidades plantean una situación problemática. Los conceptos que se introducen son muchos aunque sencillos.

- a) A partir de teoría de conjuntos y cálculo combinatorio especialmente las operaciones de unión, intersección y complemento y las de permutación y combinación se construye una función de probabilidad para eventos discretos. De las funciones continuas se escogen las que representen funciones de probabilidad para eventos continuos.
- b) A partir de las funciones de probabilidad se definen variables aleatorias, que son modelos teóricos de experimentos que después, en el capítulo de inferencia estadística se ligarán con los resultados experimentales. Es muy natural hacer su división en continuas y discretas.
- c) A partir de ambos incisos previos se define los muy intuitivos y ricos conceptos de independencia y dependencia condicional.

El teorema de probabilidad total y el de Bayes permiten entender y aplicar en ejemplos, estos conceptos claves.

Sin duda es un capítulo con muchas novedades; pero, hacer muchos ejercicios, la consulta frecuente con el asesor y la elaboración cuidadosa de las autoevaluaciones serán más que justificadas por las bases que sentarán para las siguientes unidades. Una buena medida de que se ha estudiado bien es poder reproducir todos los conocimientos en forma de resumen sin tener que revisar el proporcionado por el libro; es decir, cuando el uso frecuente de estos conceptos los ha hecho parte natural de nuestras habilidades.

En este segmento del curso es importante independizarse un poco del texto y enfatizar especialmente la comprensión de estos conceptos para las variables continuas, las más útiles en la práctica y en la que los conceptos aprendidos en Cálculo deben usarse. El resumen al final de la unidad VI puede ayudar para este objetivo así como la tabla de variables aleatorias frecuentes.

UNIDAD V

PROBABILIDAD

Hasta aquí se ha estudiado como describir una muestra, sus diversas características. Cuando se tiene certeza sobre ello, se puede proceder a buscar que hay en el fondo, que hay subyacente a estos resultados. Qué mecanismo origina la distribución que estamos observando, cuáles son las causas detrás de los resultados. Se necesita desarrollar un esquema lógico y matemático que le de sentido a nuestra intuición de que "algo" es más probable que otra cosa. Sólo a mediados de los años treinta, se elaboró un sistema axiomático para esto. Desde entonces, múltiples investigaciones teóricas y ligadas a experimentaciones en sistemas físicos, químicos, de servicio, de producción, etc. han desarrollado esta teoría que ha aportado diversos mecanismos 'aleatorios' con los que una 'probabilidad' complicada es calculada partiendo de las probabilidades de eventos elementales. En este capítulo se estudian los rudimentos de este sistema axiomático, ligándolo con los conceptos usados previamente en estadística, muestra y población, con la meta de todo estudio estadístico, la toma de decisiones informadas.

Este capítulo a pesar de su aspecto teórico es el punto de partida de modelos probabilísticos, que son muy usados en todas las áreas imaginables desde Lógica Matemática hasta el Arte pasando por las Ingenierías.

Una de las ideas más naturales al estudiar eventos es que exista entre ellas cierta dependencia o de que exista independencia. Por ejemplo, al tirar un dado la dependencia de que salgan pares y el evento que salga el 6 es "mucho". Pero la dependencia entre el evento que la UAM eleve sus colegiaturas y el nacimiento de un pez-martillo en el Mar de Cortés es inexistente.

Por otro lado, la dependencia entre el tiempo de espera a un camión y la distribución de salidas de la terminal existe pero no es tan sencilla de cuantificar. La 'probabilidad condicional' proporciona una medida numérica de la dependencia.

Estos conceptos son extraños, algo misteriosos aunque funcionan. Ha habido y sigue viva una discusión en torno a los fundamentos de la probabilidad. Pero, insistimos, sus logros son muchos y muy diversos.

Procedimiento

1. Lea del capítulo 5 las secciones 5-1 a 5-18 inclusive.
2. Resuelva los siguientes problemas de su texto: 5-3, 5-8, 5-16, 5-18, 5-19, 5-21, 5-39 y 5-41.

Autoevaluación

1. Haga un resumen exhaustivo de los temas tratados en el capítulo con sus propias palabras y sin ver el libro. Al terminar compare y complete o corrija.
2. Defina, enuncie y/o describa: la función de probabilidad, los axiomas de la probabilidad, probabilidad de espacio muestral, selección aleatoria, definición probabilística de la frecuencia relativa, eventos mutuamente excluyentes y su regla aditiva, eventos "solapados", eventos complementarios, eventos independientes, probabilidades condicionales, teorema de la probabilidad total, probabilidades a priori, probabilidades posteriores, permutaciones, combinación.
3. Considere un experimento en que una familia es seleccionada por la lotería de todas las familias que tienen exactamente 3 niños. Por ejemplo HMH denota que el primer niño que nació fue hombre, el segundo, mujer y el tercero, hombre. Hay pues 8 modos diversos

de familias HHH, HHM, HMH, HMM, MHH, MHM, MMH, MMM. (A) Qué evento corresponde al resultado compuesto de que exactamente 2 de tres niños son hombres? (B) Qué conjunto de eventos corresponde al resultado compuesto de que el primer niño nacido es mujer. (C) Suponga que cuando se observa la familia escogida en el experimento aleatorio sólo cuenta el número de niños. Es este el mismo experimento que el descrito anteriormente? ¿Cuáles son los resultados simples de este experimento? ¿Qué información se pierde al contar sólo el número de niños hombres?

4. Para la señalización de emergencia se han instalado dos indicadores que funcionan independientemente. La probabilidad de que el indicador accione durante la avería es igual a 0.95 para el primero de ellos y 0.9 para el segundo. Hallar la probabilidad de que durante la avería accione sólo un indicador. (Resultado .14)
5. Suponga que tenemos tres urnas idénticas en apariencia, conteniendo bolas de color: la urna A contiene 1 bola negra, 2 rojas y 3 verdes; la urna B contiene 2 negras, 1 roja y 1 verde; la urna C contiene 4 negras, 5 rojas y 3 verdes. Las urnas se mezclan, una de las urnas se escoge al azar y 2 bolas se extraen de ella. Si las bolas extraídas son una negra y una verde, ¿cuál es la probabilidad de que la urna B fue la escogida? (Sugerencia: use el teorema de Bayes)
6. Todas las cartas que no son de honor se han removido de un paquete de cartas, de modo que 20 cartas -los dieces, el "jack", la reina, el rey y el as de cada serie- permanece. Una carta es extraída al azar del paquete.
 - a) Dibuje el espacio muestral de este experimento.
 - b) Señale el evento A, el que un as se extraiga.
 - c) Señale el evento C, que un corazón se extraiga.

- d) Indique $A \cup C$, la unión de A con C. Cuente el espacio muestral de esta unión.
- e) Indique $A \cap C$, la intersección de A y C. Cuente los puntos de la muestra.
- f) Indique \bar{A} , el complemento de A. Cuente.
- g) Usando los conteos anteriores, encuentre $P(A \cup C)$, $P(A \cap C)$, y $P(\bar{A})$.
- h) Si se le dice que la carta seleccionada es un corazón, ¿cuál es el espacio muestral relevante reducido?, ¿cuántos puntos de la muestra se encuentran en este espacio reducido?, ¿cuál es la probabilidad^{de} que la carta seleccionada sea un as, dado que es un corazón?
- i) Usando la ley multiplicativa, encuentre $P(A \cap C)$. Compare su respuesta con la correspondiente a (g).
- j) Usando la ley aditiva, encuentre $P(A \cup C)$. Compare este resultado con su respuesta a (g).
- k) Usando complementación, encuentre $P(\bar{A})$. Compare su respuesta con la respuesta a (g).
- l) Compare $P(A|C)$ y $P(A)$. Son los eventos A y C independientes?
- m) Son los eventos A y C mutuamente exclusivos?
7. Muestre que si A y B son independientes entonces A y B^C ; B y A^C y A^C y B^C también son independientes.
8. Revise 10 problemas del texto y piense que métodos usaría para resolverlos, es decir, clasifíquelos de acuerdo con los conceptos y técnicas que les corresponden.

UNIDAD VI

DISTRIBUCIONES DE PROBABILIDAD.

Nos hemos encontrado ya con muchas distribuciones empíricas, por ejemplo la curva normal. También si tenemos un vector de variables aleatorias con elementos (x_i, y_i) o (x_1, x_2, \dots, x_n) y cada una es normal se obtiene una distribución multivariada normal. Hay razones fuertes que han sido desentrañadas por matemáticos-probabilistas como Kolmogoroff, Lévy, Feller y muchos otros, que explican la aparición tan frecuente en la naturaleza y en cualquier experimentación estadística de esta distribución y de la de Poisson. En esta unidad damos el paso crucial, entre la asignación de probabilidades a eventos, y el obtener una distribución y sus propiedades (expectación y momentos, que se relacionan con el promedio y la varianza). Es el paso de la asignación matemática de valores a la confrontación con los resultados de experimentos, es el que nos relaciona mecanismos posibles de probabilidad vistos matemáticamente, con los resultados empíricos. Los ejemplos son abundantísimos. Las primeras recompensas de tomarnos en trabajo de introducir conceptos y métodos de Cálculo a las gráficas de las distribuciones son la desigualdad de Chebyshev y algunas propiedades de los promedios y de las varianzas.

Procedimiento

1. Lea el capítulo 6, secciones: 6-1 a 6-12.
2. Resuelva los siguientes problemas de su texto: 6-6, 6-8, 6-14, 6-18, 6-23, 6-26, 6-29, 6-36, 6-37, 6-38, 6-43, 6-44, 6-49, 6-51, 6-52 y 6-60.
3. En este capítulo es especialmente importante que se haga la autoevaluación ya que incluye temas de variables aleatorias continuas que no están cubiertas en el texto.

Autoevaluación

1. La probabilidad de un evento en una distribución de frecuencia es el área bajo la curva. Cuando podemos calcular esta, por medio de integrales, (es decir, si las funciones son continuas excepto posiblemente en un número finito de puntos) entonces la función $f(x)$ por integrar, que representa a la distribución, debe tener características especiales ya que debe satisfacer los axiomas de la probabilidad.
 - a) ¿Cuáles son esas condiciones de las integrales y de las funciones?
 - b) ¿Qué dice el teorema fundamental del cálculo?
 - c) Cuánto valen $E(a) = \int_{-\infty}^{\infty} af(x)dx$, $E(bx) = \int_{-\infty}^{\infty} bx(x)dx$, $E(a+bx) = \int_{-\infty}^{\infty} (a+bx)f(x)dx$.
 - d) Cuánto valen $\text{var}(a)$, $\text{var}(bx)$ y $\text{var}(a + bx)$, sabiendo que la definición de $\text{var}(y) = \int_{-\infty}^{\infty} (y-E(y))^2 f(x)dx$.
 - e) Al volumen bajo una distribución bivariada se le asocia la probabilidad del correspondiente evento. Como los volúmenes de funciones continuas (excepto posiblemente un número finito de rectas) se pueden calcular por la integral de Riemann desarrollada en el curso de Integración, ésta impone condiciones en la función de densidad y en las integrales para poder satisfacer los axiomas de la probabilidad, análogas a las distribuciones de una sola variable aleatoria. ¿Cuáles son?
 - f) En general, una función g continua aplicada a una variable aleatoria continua x , $g(x)$, también es continua y por lo tanto, se puede calcular su integral. ¿Cuánto vale el valor esperado de $g(x)$? Ponga 3 ejemplos de g .
2. Demuestre que la siguiente función es una densidad apropiada

$$f(x) = \frac{2(b-x)}{b^2} \quad 0 \leq x \leq b$$

(Se le llama la densidad triangular.)

3. Suponga que x es una cantidad al azar con una distribución continua de probabilidad $f(x) = \frac{1}{10}$ en el intervalo $(0,10)$. Encuentre la expresión para la distribución acumulativa $F(x)$ y gráfiquela. Encuentre la mediana a partir de la gráfica de la distribución acumulativa.

Encuentre la esperanza matemática y la varianza de la distribución.

4. Aunque la desigualdad de Chebyshev no siempre es útil en la práctica (por ejemplo si $0 < h < 1$ la desigualdad de Chebyshev indica sólo que la probabilidad de la desviación no es negativa, pero esto es claro; verdad?), utilícela para estimar la probabilidad de que $|x - \mu| < 0.1$ si $\sigma_x = 0.001$. También utilice la desigualdad de μ para obtener la h y las cotas de μ , dados $P(|x - \mu| < \sigma h) \geq 0.9$ con $\sigma = 0.004$.
5. Hallar las leyes de distribución de las componentes de una magnitud aleatoria bidimensional, prefijada por la siguiente ley de distribución; es decir, encuentre las probabilidades correspondientes a los diversos valores $\{x_1, x_2, x_3\}$ y $\{y_1, y_2\}$ separadamente.

Y \ X	x_1	x_2	x_3
y_1	0.10	0.30	0.20
y_2	0.06	0.18	0.16

6. El porcentaje de alcohol en una cierta mezcla puede ser considerada con una variable aleatoria, donde x que varía entre 0 y 1 (es decir, $0 < x < 1$) tiene la siguiente densidad de probabilidad

$$f(x) = 20x^3 (1-x) \quad 0 < x < 1$$

- a) Obtenga una expresión para la F (Función de distribución acumulada) y dibuje su gráfica.
- b) Evalúe $P(x \leq 2/3)$.
- c) Suponga que el precio de venta de la mezcla depende de su contenido de alcohol. Específicamente si $\frac{1}{3} < x < \frac{2}{3}$, la mezcla se vende por C_1 pesos/litro, $x < \frac{1}{3}$ o $x > \frac{2}{3}$, se vende por C_2 pesos/litro; si el costo es de C_3 pesos/litro, encuentre la distribución de probabilidad de la utilidad neta por litro.
- 7* Suponga que un fusible tiene una vida útil X que puede ser considerada como una variable aleatoria continua con una distribución exponencial. Hay dos procesos por medio de los cuales se puede manufacturar. El proceso I conduce a una vida media de 100 horas (i.e. su parámetro es $\frac{1}{100}$) mientras que el proceso II resulta en una vida media de 150 horas (i.e. el parámetro es $\frac{1}{150}$). Suponga que el proceso II es dos veces más costoso (por fusible) que el proceso I, que cuesta C pesos por fusible. Suponga también, que si un fusible dura menos de 200 horas se incurre en una pérdida de K dólares ¿Qué proceso debe ser usado? (Res. Preferimos el proceso I, si $C > 0.13 K$).
- 8* Suponga que se tira un dado 72 veces. Dado que X es el número de veces que aparece el seis, evalúe $E(X^2)$. (Res. 154 y cuidado con tardar mucho, eso quiere decir que va mal).

*Optativos

PARZEN, Emanuel. "Tablas" en Teoría moderna de probabilidades y sus aplicaciones. Edit. Limusa-Wiley.

TABLA 6A

LAS PROBABILIDADES DE DIVERSOS EVENTOS DEFINIDOS EN LOS PROBLEMAS GENERALES DE OCUPACIÓN Y MUESTREO

La probabilidad de que		La distribución de n bolas en M urnas distinguibles		
		Sin exclusión		Con exclusión
		Bolas distinguibles	Bolas indistinguibles	Bolas ya sea distinguibles o indistinguibles
Problema de ocupación	Problema de muestreo			
I Una urna especificada contenga k bolas, donde $k \leq n$	Una bola especificada aparezca k veces en la muestra, donde $k \leq n$	$\frac{\binom{n}{k}}{M^n} (M-1)^{n-k}$	$\frac{\binom{M+n-k-2}{n-k}}{\binom{M+n-1}{n}}$	$\frac{\binom{M-1}{n-1}}{\binom{M}{n}} \quad \text{si } k = 0, 1$
II La primera urna contiene k_1 bolas; la segunda k_2 ; ...; la m -ésima k_m , donde $k_1 + k_2 + \dots + k_m = n$	En la muestra, la primera bola aparece k_1 veces; la segunda k_2 veces; ...; la m -ésima k_m veces, donde $k_1 + k_2 + \dots + k_m = n$	$\frac{\binom{n}{k_1, k_2, \dots, k_m}}{M^n}$	$\frac{1}{\binom{M+n-1}{n}}$	$\frac{1}{\binom{M}{n}} \quad \text{si } k_j \leq 1 \text{ para } j = 1, \dots, m$
III Cada una de las N urnas especificadas están ocupadas, donde $N \leq M$	Cada una de N bolas especificadas está contenida en la muestra, donde $N \leq M$	$\sum_{k=0}^N (-1)^k \binom{N}{k} \left(1 - \frac{k}{M}\right)^n$	$\frac{\binom{M-N+n-1}{n-N}}{\binom{M+n-1}{n}}$	$\sum_{k=0}^N (-1)^k \binom{N}{k} \frac{\binom{M-k}{n}}{\binom{M}{n}} = \binom{M-N}{n-N} + \binom{M}{n}$
IV Exactamente m de N urnas especificadas están ocupadas, donde $N \leq M$, $m = 0, 1, \dots, N$	Exactamente m de N bolas especificadas no están contenidas en la muestra, donde $N \leq M$, $m = 0, 1, \dots, N$	$\sum_{k=m}^M (-1)^{k-m} \binom{N}{k} \binom{k}{m} \times \left(1 - \frac{k}{M}\right)^n$ $= \binom{N}{m} \sum_{k=0}^{M-m} (-1)^k \binom{N-m}{k} \times \left(1 - \frac{m+k}{M}\right)^n$	$\frac{\binom{N}{m} \binom{M-m+n-(N-m)-1}{n-(N-m)}}{\binom{M+n-1}{n}}$ $= \frac{\binom{N}{m} \binom{M-N+n-1}{M-m-1}}{\binom{M+n-1}{n}}$	$\sum_{k=m}^M (-1)^k \binom{N}{k} \binom{k}{m} \frac{\binom{M-k}{n}}{\binom{M}{n}} = \binom{N}{m} \binom{M-N}{n-N+m} + \binom{M}{n}$
		Muestras ordenadas	Muestras desordenadas	Muestras ya sea ordenadas o desordenadas
		Con reemplazo		Sin reemplazo
La extracción de muestras de tamaño n de una urna que contiene M bolas distinguibles				

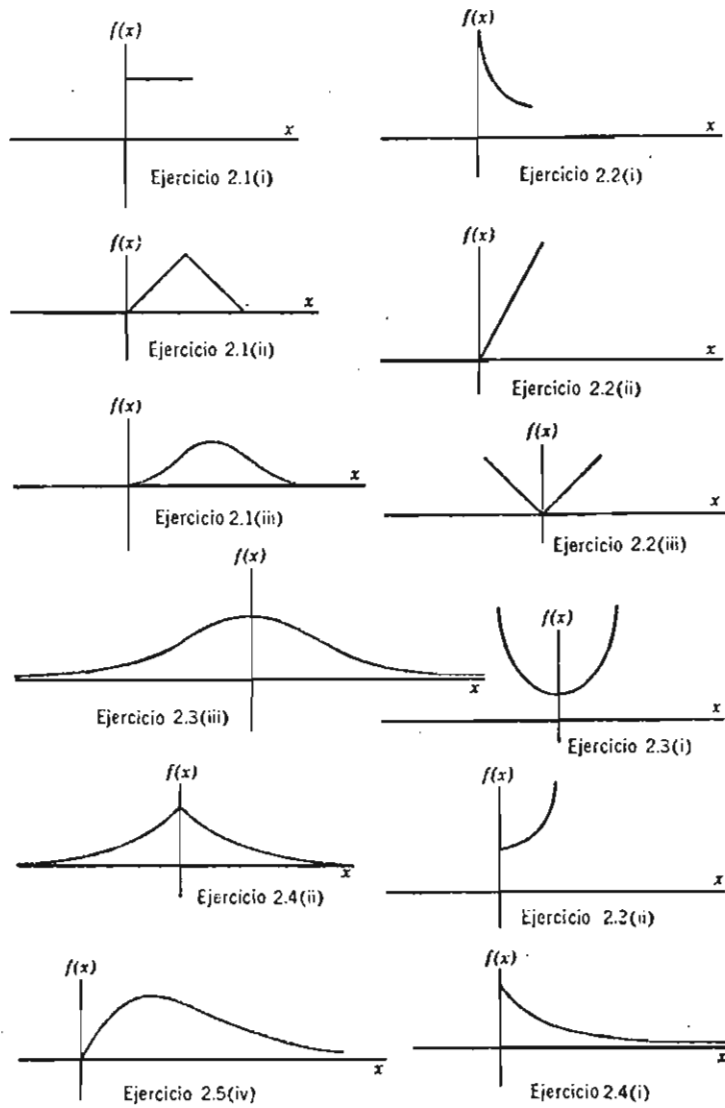


Fig. 2A. Gráficas de las funciones de densidad de probabilidades que se dan en los ejercicios indicados.

TABLA 5A

Número de maneras de distribuir n bolas en M urnas distinguibles			
Bolas distribuidas	Bolas distinguibles	Bolas indistinguibles	
Sin exclusión	M^n Estadísticas de Maxwell-Boltzmann	$\binom{M+n-1}{n}$ Estadísticas de Bose-Einstein	Con reemplazo
Con exclusión	$(M)_n$	$\binom{M}{n}$ Estadísticas de Fermi-Dirac	Sin reemplazo
	Muestras ordenadas	Muestras desordenadas	Muestras extraídas
Número de maneras de extraer muestras de tamaño n de una urna que contiene M bolas distinguibles			

UNIDAD VII

MODELOS PROBABILÍSTICOS ESPECIALES

Habiendo comprendido los rudimentos de la teoría de la probabilidad, nos interesa adentrarnos en la estructura de modelos útiles y comunes que nos sirvan para aproximar las distribuciones estadísticas que se nos presentan. Tanto para variables discretas como para las continuas deseamos tener un modo standard, tabulado, de calcular, modulando adecuadamente los parámetros, las probabilidades de eventos definidos.

Estos modelos presentan interrelaciones, por ejemplo la distribución normal es el "límite" de la binomial, la normal multivariada el "límite" de la multinomial. La binomial aproxima la hipergeométrica y la Poisson aproxima a la binomial. Todo esto es de utilidad en determinadas instancias de control de calidad para procesos industriales, por ejemplo. También lo es para la demostración de los teoremas estructurales de la teoría de Probabilidad y de Procesos Estocásticos.

Procedimiento

1. Lea el capítulo 7, secciones 7-1 a 7-13.
2. Resuelve los siguientes problemas: 7-3, 7-5, 7-6, 7-9, 7-16, 7-17, 7-22, 7-36, 7-39, 7-46, 7-53, 7-55, 7-56, 7-61, 7-68.

Autoevaluación

1. Defina y/o describa los siguientes conceptos: modelos probabilísticos, modelo de Bernoulli, modelo binomial, función probabilística, binomial, distribución de una variable binomial, función de probabilidad de una variable multinomial, función probabilística de una variable hipergeométrica, modelo de probabili-

dad uniforme o rectangular, modelo de Poisson, distribución gamma, modelo normal general y estandarizado.

2. a)* Una muestra aleatoria simple se define como una muestra extraída bajo la condición de que todas las muestras posibles tengan la misma probabilidad de ser seleccionadas. Para este tipo de muestras cada elemento de la población es igualmente probable de ser incluida en una de ellas. Suponga que un plan de muestreo es diseñado de modo que la población de la muestra de tamaño $N = 100$ es dividido en 10 estratos y que un elemento unitario de cada estrato se selecciona al azar, para una muestra de tamaño 10.

¿Es ésta una muestra aleatoria de acuerdo con la definición dada? ¿Es cualquier elemento de la población igualmente probable de ser elegido? ¿Es la probabilidad de escoger un elemento cualquiera de la población es la misma?

- b) En las últimas veinte semanas, la distribución de accidentes que ocasionan pérdida de tiempo ha sido de nueve semanas sin accidentes, 5 semanas con un accidente, cuatro semanas con 2 accidentes, una semana con tres accidentes y una semana con cuatro accidentes.

i) ¿Cuál es la razón media de accidentes por semana?

ii) Para esta razón media, ¿cuál es la probabilidad de que haya 0,1,2,3, y 4 o más accidentes en una semana dada? Use el modelo de Poisson.

iii) Usando el modelo de Poisson, ¿cuál es la probabilidad de que no habrá accidentes en 2 semanas?

- c) Para una variable aleatoria normal X , con $\mu_x = 10$ y $\sigma_x^2 = 225$ encuentre las siguientes probabilidades.

*Optativo

- | | |
|---------------------------|--|
| a) $P(x \leq 10)$ | f) $P(-30 \leq x \leq 5)$ |
| b) $P(x \geq 10)$ | g) $P(5 \leq x \leq 90)$ |
| c) $P(x \leq 0)$ | h) $P(x \leq -5 \text{ o } x \geq 25)$ |
| d) $P(0 \leq x \leq 10)$ | i) $P(x \leq 2 \text{ des. } x)$ |
| e) $P(15 \leq x \leq 50)$ | j) $P(-D.S. \leq x \leq D.S.)$ |
- (D.S. = desviación standard).
= des. x)

3. a) ¿Cuál es la probabilidad de que una familia elegida al azar tenga un ingreso de más de \$11,000.00 si la media es de \$10,000.00 y la desviación standard es de \$2,500.00? y la distribución de ingresos es normal?
- b) Menor que \$8,500.00 .
- c) Entre \$11,500.00 y \$12,000.00?
- d) Entre \$9,000.00 y \$11,000.00?
- e) ¿Qué intervalo, centrado en el ingreso promedio familiar, contendrá al 50 por ciento de las familias?

UNIDADES VIII, IX, X

INFERENCIA ESTADÍSTICA

Después de haber visto tanto material, tantos nuevos conceptos al llegar a este nivel, se pregunta uno, ¡cómo! ¿aún hay más? ¡Si ya no puede existir algo diferente a lo que ya estudiamos!

Claro que existe y mucho e importante.

Qué es lo que se ha visto: en las primeras 4 unidades una introducción y una descripción elemental de experimentos estadísticos, hasta incluir \bar{X} y S_x^2 .

En las siguientes tres unidades, todo el instrumental matemático - al servicio de nuestra habilidad de modelación, hasta incluir μ_x y σ_x^2 (o μ_x y σ_x^2 aproximadas al usar los modelos).

Nos falta lo esencial, la síntesis de las dos tesis anteriores (la parte puramente empírica y la parte puramente teórica) y aquí es donde surgen los teoremas mas profundos, las aplicaciones mas fecundas.

El problema a resolver es el ligar los múltiples histogramas que pueden resultar en un mismo experimento, las múltiples \bar{X} y S_x^2 que resultan de muestras diferentes de un mismo experimento, las diversas trayectorias que puede tener un objeto moviéndose al azar con las características esenciales, poblacionales, teóricas posibles.

Es decir dada una muestra qué se puede y con qué seguridad se puede concluir algo sobre el fenómeno en sí.

Dos muestras diferentes corresponden a la misma población? ¿Con qué margen de seguridad puede afirmarse esto? El promedio muestral qué me permite inferir respecto a la media de la población? Está este en un intervalo que pueda yo precisar?

Aún más dramática situación: si debo concluir a partir de mi humilde muestra algo respecto a la población. Qué manipulaciones algebraicas de la media y la varianza muestral me garantizan algo respecto a las poblacionales? Cómo puedo asegurarlo.

Y se es consciente que por muy buenas que sean mis matemáticas, la muestra por limitada y barata, puede conducirme a concluir cosas erróneas. Puedo de algún modo limitar la magnitud de mis errores, hacer a estos menos probables?

El problema de como muestrear se introduce pero no se cubre en realidad.

UNIDAD VIII.

TEORÍA DEL MUESTREO

En cualquier investigación real estamos interesados en poder hacer afirmaciones globales, de la población, a partir de los datos de la muestra; poder estimar parámetros de la población. Para estimar bien, se necesita tomar una buena muestra. Una muestra que sea representativa. Se pueden obtener muestras representativas mediante asignación de números al azar a los elementos de la población y tomando una lista de ellos o extrayéndolos de urnas como en la lotería; pueden extraerse muestras con reemplazamiento o sin él.

Una población puede pensarse dividida en muestras y calcular a cada una de estas su media, su mediana, su varianza. Se puede construir un nuevo espacio de probabilidad con cada muestra como nuevo elemento y hablar así de la distribución de promedios muestrales, de varianzas muestrales etc. Los métodos difieren según que las muestras sean grandes o pequeñas en tamaño.

Procedimiento

1. Leer el capítulo 8, secciones: 1-11.

Hacer los problemas: 8-15, 8-18, 8-19, 8-21, 8-23, 8-37, 8-43, 8-48, 8-53, 8-55.

2893128

2. Haga un resumen exhaustivo de los temas tratados en el capítulo, con sus propias palabras y sin ver el libro. Al terminar compare y complete o corrija.
3. Defina, enuncie y/o describa: teorema de Khintchine, ley de Bernoulli, ley débil y ley fuerte de los grandes números, teorema central del límite, muestreo, distribución de probabilidad

de una estadística, distribución muestral del número de éxitos, factor de corrección por continuidad, factor de corrección de población finita, diferencia de proporciones muestrales, diferencia de medias muestrales.

Autoevaluación

1. Un banquero que maneja valores está interesado en encontrar la cantidad de dinero disponible de sus clientes para propósitos de inversión. Sin saberlo él, el promedio de dinero de inversión por cliente es \$5,000, con una desviación standard de \$2,000. Si el banquero toma una muestra de 100 clientes, ¿cuál es la probabilidad de que el total disponible a los miembros de esta muestra sea más de \$550,000.00? ¿Cuál es la probabilidad de que la cantidad promedio por cliente de la muestra sea menos de \$5,100.00?

El banquero sabe que la distribución de dinero de inversión por cliente no está distribuida normalmente. La distribución tiene una muy larga cola derecha; hay muchos clientes con nada de dinero para inversión. ¿Afecta esta asimetría a sus respuestas anteriores? ¿Porqué?

2. Suponga que 5 por ciento de la fuerza de trabajo en un pueblo A está desempleada y que 8 por ciento de la fuerza de trabajo en B está desempleada. Una muestra aleatoria de la fuerza de trabajo es tomada en cada pueblo. ¿Cuál es la probabilidad de que los resultados de la muestra indiquen que el pueblo A tiene desempleo más alto que B?
3. El diámetro promedio de una correa es de 10 mm y $\sigma = 0.25$ mm. ¿Cuál es la probabilidad de seleccionar una muestra aleatoria de tamaño $n = 25$ correas que tenga una media muestral de 9.2 mm. o menor?

4. Se ha demostrado, por datos que se han tomado, que el porcentaje de transistores dañados en un embarque es del 20%. ¿Cuál es la probabilidad de seleccionar una muestra de tamaño 100 con una proporción muestral del 25% o más? Suponga que el factor de corrección para población finita es $= 1$.
5. Se sabe que un promedio del 7% de los animales domésticos menores de un año mueren en cierta región. Nacieron 100 cachorros durante un cierto período de tiempo. ¿Cuál es la probabilidad de que 10 o más de los cachorros mueran durante su primer año?

UNIDAD IX.

ESTIMACIÓN

El conocimiento preciso de los parámetros de una población no es accesible generalmente. Sin embargo en la práctica laboral se hace necesario tomar decisiones del mejor modo posible en base a los datos recabados. Para esto se hacen estimaciones de magnitudes estadísticas sea en forma exacta "puntual", o dando un intervalo en que se tenga una certeza aproximada de que el parámetro se encuentra. Hay diversos 'estimadores' para calcular estos parámetros y se definen una serie de requisitos naturales que es deseable que satisfagan como insesgabilidad, suficiencia, consistencia y eficiencia. Esta parte del curso tiene especial utilidad práctica y es parte de las innovaciones que surgieron en los años treinta de este siglo y que indujeron a considerar a la estadística, más y más, como una teoría matemática de la decisión como se puso de manifiesto con los trabajos de Neyman - Pearson, de Wald (después) en base a experimentaciones industriales, etc.

La teoría de juegos ideada y desarrollada por Von Neumann y aplicada junto con Morgenstern a la economía (en los años de la segunda guerra mundial) fue una búsqueda paralela de toma de decisiones óptimas en situaciones de conflictos reales, modeladas con matemáticas y probabilidad.

Procedimiento

1. Lea el capítulo 9, secciones 1-11.
2. Resuelva los siguientes problemas: 9-10, 9-14, 9-15, 9-19, 9-28, 9-36, 9-37, 9-38.
3. Defina y/o describa los siguientes conceptos: inferencias estadísticas, estimación, docimasia, estimación puntual, parámetro

insesgado, parámetro consistente, parámetro más eficiente que otro, estimador suficiente, función de máxima verosimilitud, estimación de máxima verosimilitud, estimación por intervalo, intervalo de confianza, multiplicador de confianza, límite inferior de confianza y límite superior de confianza.

Autoevaluación

1. Suponga que un objeto es medido con dos instrumentos diferentes. Sean L_1 y L_2 las medidas obtenidas. Si ambos instrumentos están calibrados correctamente podemos suponer que $E(L_1) = E(L_2) = L$, la longitud correcta. Sin embargo, la exactitud de los instrumentos no es necesariamente la misma. Si medimos la exactitud en términos de la varianza, entonces $V(L_1) \neq V(L_2)$. Si usamos la combinación lineal $Z = aL_1 + (1 - a)L_2$ como nuestra estimación de L , tenemos inmediatamente que $E(Z) = L$. Esto es, Z es una estimación no-sesgada de L . ¿Para que elección de a , $0 < a < 1$ es la varianza de Z un mínimo?

2. Piense el siguiente experimento: Se está tomando una muestra de una población con parámetros: media = 5 y varianza = 10. Para un gran número de veces $Q = 1,000,000$ se toma una muestra de tamaño 16 de la población. Para cada muestra de 16 observaciones se señala la media (un estimador de la media de la población) la varianza s^2 de la muestra con divisor n (un estimador de la varianza de la población) y la varianza s'^2 de la muestra con divisor $(n-1)$ (también un estimador de varianza de la población). Después de extraer estas Q muestras, se encuentra el promedio aritmético de todos los valores de \bar{x} , la media aritmética de todos los valores de s^2 , y la media aritmética de todos los valores s'^2 .
 - a) Prediga la media aritmética de los valores de \bar{x} .
 - b) Prediga la media aritmética de los valores de s^2 .

- c) Prediga la media aritmética de los valores de s'^2 .
 - d) Las predicciones anteriores se realizarán precisamente o só lo provisionalmente.
 - e) Qué cambiaría en el experimento para hacer sus predicciones más cercanas?
 - f) Qué cambiaría en el experimento conceptual para reducir en diferencia a las respuestas en la parte b y c?
-
- h) Indique las partes de esta cuestión que ilustran las propie dades de estimadores no sesgados.
 - i) Cómo ilustraría la propiedad de insesgabilidad asintótica.
 - j) Indique la parte de la pregunta que ilustra la consistencia.
3. Un analista de operaciones de una compañía aérea estaba investigando el número de minutos de los vuelos que llegaron después de la hora de llegada publicada. Seleccionó al azar 600 llegadas; de éstas, 375 fueron puntuales y fueron descartadas de la muestra porque no vinieron de la población de llegadas tardías. De las llegadas restantes, las siguientes estadísticas se calcularon:
- $$\bar{x} = 18 \text{ minutos}$$
- $$s = 9 \text{ minutos}$$
- a) ¿Cuál sería la idea sobre la asimetría de la población de los minutos posteriores a la llegada correcta de los vuelos tardíos?

- b) Defienda el uso de la población normal para encontrar el multiplicador de confianza, aún cuando la población no es normal.
 - c) Defienda el uso de la desviación standard de la muestra en vez de la desviación standard de la población.
 - d) Encuentre el intervalo de confianza de 95% para los minutos de llegadas tardías.
4. Un bioquímico desea averiguar el promedio de contenido vitamínico de un alimento. El error debe estar a más o menos 0.5 mg. Con un nivel de confianza del 95%. ¿Qué tamaño debe tener la muestra? Suponga que la desviación standard es de $s = 2$ mg.

UNIDAD X

DOCIMASÍA ESTADÍSTICA

Es frecuente la necesidad de tomar decisiones y de tener una medida de la bondad de esta decisión. En esta unidad se precisan estos conceptos haciéndolos accesibles al cálculo numérico. Hay dos tipos de errores que se pueden cometer al hacer una hipótesis en cuanto al valor de una estadística:

Que la hipótesis se tome como verdadera siendo falsa en realidad (error tipo I) o que se tome como falsa siendo verdadera (error tipo II). Se pretende obtener un proceso standard para evaluar estos parámetros, que consta de las etapas siguientes:

Hipótesis, nivel de significación, docimasia estadística, regla de decisión, cálculos y decisiones. Una vez más, el planteo del problema es enfatizado.

Procedimiento

1. Leer el capítulo 10, secciones: 10-1 a 10-6.
2. Hacer los problemas: 10-11, 10-13, 10-14, 10-17, 10-25, 10-28.
3. Haga un resumen exhaustivo de los temas tratados en el capítulo con sus propias palabras y sin ver el libro. Al terminar compare y complete o corrija.
4. Defina y/o describa: hipótesis nula, hipótesis alternativa, criterio de decisión para aceptar, región de rechazo, región de aceptación, error del primer tipo, nivel de significación, error del tipo II, función de potencia de la docimasia, docima

sia de una sola cola, docimasia bilateral, hipótesis compuesta, región crítica o de rechazo, región de aceptación o de no rechazo.

Autoevaluación

1. Una oficina de control de calidad debe tomar una decisión sobre un grupo de cinturones de seguridad. Las especificaciones para los asientos requieren de una longitud de 30 pulgadas. La experiencia previa con el proveedor ha revelado que la desviación standard de la longitud de los cinturones es de 2 pulgadas y que las longitudes están distribuidas normalmente. La oficina de control de calidad cree que los cinturones son satisfactorios; sin embargo, si la longitud promedio es muy pequeña o muy grande, el cinturón será rechazado. Los cinturones no serán rechazados sin aportar la evidencia. Se determina que 16 será el tamaño de la muestra.
 - a) ¿Cuáles son las hipótesis nula y alternativa?
 - b) Si la probabilidad de un error de tipo I ha de ser de 0.01, ¿cuál es el nivel de significancia?
 - c) ¿Qué estadística se debería usar?
 - d) ¿Qué criterios de decisión se deben usar, en términos de la estadística de prueba, \bar{x} ?
 - e) ¿Qué criterios de decisión se deberían usar, en términos de la estadística z ?
 - f) Si la longitud promedio de 16 cinturones es de 29.74 pulgadas, ¿qué acción se debería tomar? Responda usando primero a la \bar{x} y después a la z como la estadística de prueba.

2. Un vendedor de la compañía de teléfonos ha afirmado que 35% de las llamadas de larga distancia a los clientes resultará en la recepción de un orden. Se le pide que diseñe un experimento con 64 llamadas para demostrar su hipótesis. Han acordado en un 10% de nivel de significancia para una prueba de una cola.
 - a) ¿Cómo debería el vendedor presentar su hipótesis para que su afirmación tenga una mejor oportunidad de utilizar la prueba?
 - b) ¿Cómo se debería diseñar el test si se quisiera ser especialmente cuidadoso para evitar el error de usar larga distancia cuando la productividad de las llamadas es menor que 35%. Su acción sería usar larga distancia si el vendedor "gana" la discusión.
3. Se indica la variabilidad de fabricación de cierta pieza de relojes de pared por una desviación standard de 0.02 pulgadas, que, de acuerdo con la experiencia permanece estable. Se especifica que esta pieza tiene un diámetro externo de $2.06 \pm .03$ pulgadas. Cuando no se cumple esta especificación, se considera que el proceso de producción está fuera de control y se debe detener para su ajuste. Si $\alpha = 0.01$ y $\beta = 0.10$. ¿Cuál es el tamaño de muestra que se requiere con el propósito de verificar si el proceso está o no bajo control?, ¿cuáles son los valores críticos para esta regla de decisión?

PRIMERA UNIDAD DE REVISIÓN

En el estudio de la Probabilidad y Estadística, las necesidades computacionales, pueden inducir a despreciar la ordenada y potente estructura lógica. Esta unidad es para revisar los diversos conceptos estudiados y tratar de relacionarlos, especialmente los que surgen del experimento con los conceptos teóricos de la probabilidad.

El independizarse de los diversos algoritmos y verlos como lo que son, herramientas para poder juzgar con la conciencia tranquila, es importante, pero no siempre es un paso que los alumnos de cursos elementales dan.

El darse cuenta que estos algoritmos pueden deformarse, como se ilustra abundantemente en el libro de Huff, creemos que ayudará a usarlos en forma cuidadosa e inteligente.

En el curso básicamente se ha estudiado estadística descriptiva. Así será aún por el resto del curso. Sin embargo la 'descripción' puede ser utilizada para la 'predicción' y esto, la parte más dinámica de la probabilidad, se usa ampliamente en Física y Comunicaciones, para poner sólo dos ejemplos.

Lo que hacemos en este curso es necesario para entender las partes más finas de las aplicaciones. Curiosamente, algunos de los temas más abstractos y difíciles de las matemáticas como la teoría de los números y la teoría de la programación, requieren de muchos de estos conceptos.

Procedimiento

1. Lea los capítulos 9 y 10 del Huff.
2. Comprenda los problemas y proyectos teóricos de autoevaluación

de las precedentes unidades. Por un lado debe diferenciar los casos de variables discretas y las de continuas y las formas matemáticas que presentan sus diversas propiedades como media, varianza y momentos. Diferenciar también las formas de estos parámetros para muestras y para poblaciones teóricas, así como entre distribuciones univaluadas y multivaluadas.

3. Haga una lista de los principales algoritmos (métodos de cálculo) de los diversos parámetros estudiados para las variables aleatorias estudiadas.
4. Recuerde las propiedades de la integral que son utilizadas para la evaluación de eventos asociados a variables aleatorias continuas.
5. Revise el Huff y haga una lista sucinta de los principales vicios que menciona.

Autoevaluación

1. Diga si es falso o verdadero y por qué.
 - a) El promedio aritmético de una población nunca será mayor que el agregado para esta población.
 - b) Al efectuar una investigación para determinar el ingreso promedio en un distrito, las unidades elementales en la población que va a ser estudiada serán los hombres mayores de 21 años que son jefes de hogar y que residen en el distrito.
 - c) Una curva de frecuencia acumulativa construida en la base de "menor que" siempre crecerá a menos que intervalos desiguales obliguen a la curva a caer.
 - d) La curva de frecuencia acumulativa de un conjunto de datos

cuyo histograma tiene todas las barras de igual tamaño sería una línea recta.

- e) La desviación promedio de una muestra con una media cero será cero, si el rango es cero.
 - f) Una distribución simétrica usualmente tiene el promedio ligeramente mayor que la mediana.
 - g) Una variable aleatoria puede ser pensada como una regla para trasladar los resultados de un experimento del espacio muestral al conjunto de los números reales.
 - h) El valor de una función de distribución acumulativa para una variable aleatoria puede ser hasta dos para algunas poblaciones.
2. Una tienda está investigando el número de televisores vendidos por veinte vendedores el año pasado. El número de televisores vendidos por cada vendedor fueron: 26, 34, 21, 35, 31, 48, 49, 77, 71, 67, 69, 68, 50, 55, 51, 53, 58, 62, 59 y 46.
- a) ¿Cuál es el rango de las observaciones?
 - b) Usando un intervalo de clase de 10 y un límite para la clase inferior de 20, cuál es el número de clases necesarias? son estas clases, intervalos de clase y límites, convenientes? ¿Qué tan cercanamente se conforma este conjunto de clases a la guía aproximada dada por la regla de Sturges?
 - c) Encuentre el promedio, la mediana y la moda.
 - d) Que suposición se hace para aplicar la fórmula

$$\text{Med} = I + \left(\frac{n/2 - \sum f_1}{f_{\text{med}}} \right) i$$

dado en el texto, para calcular la mediana.

3. a) Usando la fórmula 4.14, encuentre la desviación standard para los datos del problema 2.
- b) El vendedor que vendió más fue el de los 77 televisores. A cuántas desviaciones estaba del promedio.
- c) Encuentre el primero, segundo y tercer cuartil de los datos del problema 2.
4. La siguiente tabla da, los porcentajes de estudiantes que usan anteojos, en cada clase de la Universidad.

	PORCENTAJE QUE USAN LENTES	NO. TOTAL EN LA CLASE
1º año	30	600
2º año	25	700
3º año	20	800
4º año	15	900

- a) Se selecciona al azar un estudiante del conjunto de éstos. ¿Cuál es la probabilidad de que use anteojos? (Use el teorema de probabilidad total)
- b) ¿Cuál es la probabilidad de que sea estudiante de 1º ó 2º -- año dado que usa anteojos? (Use el teorema de Bayes)
5. a) Un experimento consiste en tirar dos dados simultáneamente. Sea la variable aleatoria X la suma de los resultados. Exhiba la distribución de probabilidad de X en forma tabular.
- b) De la variable aleatoria X descrita en el problema 2 de la sexta unidad, encuentre EX , la varianza y la desviación standard. Res. $(\frac{5}{3}, \frac{b^2}{18}, \frac{b}{3\sqrt{2}})$
- c) Calcule la función de distribución acumulativa para esta X triangular $(F(x) = \frac{2}{b^2} (bx - \frac{x^2}{2}) \quad 0 \leq x \leq b, F(x) = 1 \text{ para } x > b; F(x) = 0 \text{ para } x < 0)$.

SEGUNDA UNIDAD DE REVISIÓN

2893128

Los métodos estudiados en el curso aunque muy extensos y en ocasiones tediosos, constituyen el principio, las bases sobre las que se construyen métodos para resolver problemas de otro tipo. Desgraciadamente quedan muchos temas por cubrir que son útiles y de uso frecuente.

Por ejemplo, en ciencias sociales una herramienta muy utilizada es el análisis de varianza: que "causas" originan la varianza de alguna característica; 'qué tanto' de la varianza se debe a una variable dada. Ya establecido que una variable depende de otras en forma apreciable, que función lineal, cuadrática o bien definida se ajusta mejor, (regresión lineal, cuadrática, etc.).

Cómo sabemos qué factores son los importantes al estudiar un fenómeno cualquiera; qué variables están ligadas y cuáles desligadas entre sí (análisis de correlación simple ó múltiple).

Un buen modo de terminar el curso es darse una idea esquemática de que más hay en probabilidad y estadística. Para estadística consúltese el artículo adjunto 'Some classes of Problems in Statistical Methods'. Para probabilidad, el artículo de 'Applications to Operations Research' y 'Roles of Probability in Engineering'.

Otro tema que se ha desarrollado en buena medida por las intensas presiones de tecnólogos e ingenieros es el estudio de procesos estocásticos que se usan frecuentemente en comunicaciones y meteorología, en física y en electrónica y, en general, en sistemas.

Hay algunos modelos (ej. Cadenas de Markov) que parten del concepto de probabilidad condicional y que permiten obtener resultados útiles en forma bastante sencilla por ejemplo en economía con la matriz de insumo-producto, en procesos de aprendizaje, etc.

Regresando al curso vemos que se cubren temas que permiten cuantificar probabilidades y describir poblaciones mediante muestras. Se enunciaron algunas de las principales leyes de la probabilidad clásica como el teorema del límite central, las leyes de los grandes números, ambas piezas clave para poder hablar de muestreo; sin embargo estas afirmaciones y otras que las generalizan requieren de un aparato matemático, creado a principios de este siglo que amplía y fortalece al concepto de Integración de Riemann-Stieltjes. Esta nueva integral (de Lebesgue) permite fundamentar con mucha precisión los conceptos de variable aleatoria, de sucesiones de variables aleatorias y de problemas de convergencia de éstas. Además, involucra el uso de transformadas de Fourier y operaciones entre éstas que son un arma muy útil para calcular las propiedades de las distribuciones.

Esta unidad finaliza el curso y, la oportunidad de organizar el material, superar deficiencias y volverlo práctico para su uso real es inmejorable. Se han usado muchos métodos; la meta de esta unidad es desarrollar pericia para usarlos y distinguirlos así como el profundizar en su sustento teórico. Si se logra integrar la conciencia crítica, la habilidad técnica y la claridad teórica los fines de este curso se habrán alcanzado.

Procedimiento

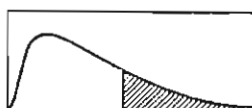
1. Comprenda los problemas y proyectos teóricos de autoevaluación de las unidades precedentes (con énfasis en las últimas 4 pero incluyendo las anteriores).
2. Haga una lista de los principales algoritmos utilizados en las unidades de estimación y docimasia, así como en las aproximaciones y desarrollo de los modelos probabilísticos y los cálculos de teoría de muestreo. Rehaga una lista de definiciones y de enunciados de los principales resultados teóricos.

TABLAS

Y. L. Chou

Cuadro A-VIII

LA DISTRIBUCIÓN CHI CUADRADO



df	Probabilidad de que el valor de chi cuadrado sea excedido							
	.995	.990	.975	.950	.050	.025	.010	.005
1	---	---	---	.004	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	90.53	95.02	100.43	104.22
80	51.17	53.54	57.15	60.39	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	124.34	129.56	135.81	140.17

Cuadro A-IX

LA DISTRIBUCIÓN F

Puntos superiores al 10 por 100



$\frac{1}{\sigma}$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	100	∞
1	30.80	49.50	63.50	65.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.63	9.00	9.16	9.21	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	6.54	6.46	6.39	6.34	6.31	6.28	6.27	6.25	6.24	6.23	6.22	6.20	6.18	6.18	6.17	6.16	6.15	6.14	6.13
4	4.64	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.00	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.88	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.84	1.82	1.79	1.76
16	3.03	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.00	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	2.97	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.90	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.87	2.60	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.86	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.85	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.84	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.83	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.82	2.53	2.32	2.18	2.09	2.03	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.81	2.52	2.31	2.17	2.08	2.02	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.80	2.51	2.30	2.17	2.07	2.01	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.80	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.80	2.50	2.28	2.15	2.05	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.80	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.80	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.76	2.36	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Cuadro A-IX

(continuación)

Puntos superiores al 5 por 100

δ_1	δ_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	181.4	192.5	215.7	224.6	230.2	234.0	236.8	238.9	240.6	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.48	19.47	19.48	19.49	19.50	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.68	8.64	8.62	8.59	8.57	8.55	8.53	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.88	5.80	5.77	5.75	5.72	5.69	5.66	5.63	
5	8.61	6.79	6.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.58	4.53	4.50	4.46	4.43	4.40	4.36	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.08	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	
8	6.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.16	3.12	3.08	3.04	3.01	2.97	2.93	
9	5.12	4.28	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	
14	4.60	3.74	3.34	3.11	2.96	2.86	2.78	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	
15	4.54	3.68	3.28	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	
20	4.36	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25	
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	

Cuadro A-IX
(continuación)
Puntos superiores al 1 por 100

δ_1 δ_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4998.5	6103	5625	5784	5859	5928	5982	6022	6058	6106	6157	6209	6255	6281	6297	6313	6339	6366
2	86.50	89.00	92.17	95.25	98.20	99.33	99.76	99.97	99.99	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	28.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	19.00	16.69	15.98	15.32	16.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.28	13.27	12.06	11.39	10.97	10.67	10.48	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.16	8.75	8.47	8.28	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.53	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.66	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.85	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.55	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.18	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.68	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.16	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.68	4.78	4.28	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.65	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.66	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.28	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.93	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.96	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.60	1.47	1.32	1.00

Fuente: E. S. Pearson y H. O. Hartley, Biometrika Tables for Statisticians, volumen 1. Cuadro 18.

Cuadro A-XI

$$\text{Valores de } z = \left(\frac{1}{2}\right) \ln \left(\frac{1+r}{1-r}\right)^*$$

(Para los valores negativos de r coloque un signo menos delante de los números del cuadro.)

r	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.00000	.01000	.02000	.03001	.04002	.05004	.06007	.07012	.08017	.09024
0.1	.10034	.11045	.12058	.13074	.14093	.15114	.16139	.17167	.18198	.19234
0.2	.20273	.21317	.22360	.23419	.24477	.25541	.26611	.27686	.28768	.29857
0.3	.30952	.32055	.33165	.34283	.35409	.36544	.37689	.38842	.40000	.41180
0.4	.42365	.43561	.44769	.45990	.47223	.48470	.49731	.51007	.52298	.53606
0.5	.54931	.56273	.57634	.59014	.60415	.61839	.63283	.64752	.66246	.67767
0.6	.69315	.70892	.72500	.74142	.75817	.77530	.79281	.81074	.82911	.84795
0.7	.86730	.88718	.90764	.92873	.95048	.97205	.99621	1.02033	1.04537	1.07143
0.8	1.09861	1.12703	1.15682	1.18813	1.22117	1.25615	1.29334	1.33308	1.37577	1.42192
0.9	1.47222	1.52752	1.58902	1.65839	1.73805	1.83178	1.94591	2.09229	2.29756	2.64605

* Extractada con permiso de Wilfrid J. Dixon y Frank J. Massey, Jr., *Introduction to Statistical Analysis*, 2a. edición, p. 468. McGraw-Hill Book Company, New York, 1957.

Cuadro A-XII

PUNTOS DE SIGNIFICACIÓN DE 5 POR 100 Y DE 1 POR 100 PARA EL COEFICIENTE DE CORRELACIÓN SERIAL (DEFINICIÓN CIRCULAR)*

n	Cola positiva		Cola negativa	
	5%	1%	5%	1%
5	0.253	0.297	-0.753	-0.798
6	0.345	0.447	0.708	0.863
7	0.370	0.510	0.674	0.799
8	0.371	0.531	0.625	0.764
9	0.366	0.533	0.593	0.737
10	0.360	0.525	0.564	0.705
11	0.353	0.515	0.530	0.679
12	0.348	0.505	0.516	0.655
13	0.341	0.495	0.497	0.634
14	0.335	0.485	0.479	0.615
15	0.328	0.475	0.462	0.597
20	0.299	0.432	0.390	0.524
25	0.276	0.398	0.356	0.473
30	0.257	0.370	0.325	0.433
35	0.242	0.347	0.300	0.401
40	0.229	0.329	0.279	0.376
45	0.218	0.314	0.262	0.356
50	0.208	0.301	0.248	0.339
55	0.199	0.289	0.236	0.324
60	0.191	0.278	0.225	0.310
65	0.184	0.268	0.216	0.298
70	0.178	0.259	0.207	0.287
75	0.173	0.250	-0.199	-0.276

Para los valores de n por encima de 75, utilice las siguientes fórmulas para determinar los puntos de significación:

$$\text{Para el nivel de significación de 5 por 100: } \frac{-1 \pm 1.645 \sqrt{n-2}}{n-1}$$

$$\text{Para el nivel de significación de 1 por 100: } \frac{-1 \pm 2.326 \sqrt{n-2}}{n-1}$$

R. L. Anderson, "Distribution of the serial correlation coefficient," *Annals of Mathematical Statistics*, 13, No. 1, 1942, pp. 1-13.

* Reproducido con permiso de los editores.

Cuadro A-XIII

PUNTOS DE SIGNIFICACIÓN DE 5 POR 100 Y DE 1 POR 100 PARA EL COEFICIENTE ENTRE LA DIFERENCIA
MEDIA CUADRÁTICA SUCESIVA Y LA VARIANZA*

Valores de $\frac{\delta^2}{s^2}$ para los diferentes niveles de significación

Valores de k					Valores de k'				
n	$P = .01$	$P = .05$	$P = .95$	$P = .99$	n	$P = .01$	$P = .05$	$P = .95$	$P = .99$
4	.8341	1.0406	4.2927	4.4992	31	1.2469	1.4746	2.6587	2.8864
5	.6724	1.0255	3.9745	4.3276	32	1.2570	1.4817	2.6473	2.8720
6	.6738	1.0682	3.7318	4.1262	33	1.2667	1.4885	2.6365	2.8583
7	.7163	1.0919	3.5748	3.9504	34	1.2761	1.4951	2.6262	2.8451
8	.7575	1.1228	3.4486	3.8139	35	1.2852	1.5014	2.6163	2.8324
9	.7974	1.1524	3.3476	3.7025	36	1.2940	1.5075	2.6068	2.8202
10	.8353	1.1803	3.2642	3.6091	37	1.3025	1.5135	2.5977	2.8085
11	.8706	1.2062	3.1938	3.5294	38	1.3108	1.5193	2.5889	2.7973
12	.9033	1.2301	3.1335	3.4603	39	1.3188	1.5249	2.5804	2.7865
13	.9336	1.2521	3.0812	3.3996	40	1.3266	1.5304	2.5722	2.7760
14	.9618	1.2725	3.0352	3.3458	41	1.3342	1.5357	2.5643	2.7658
15	.9880	1.2914	2.9943	3.2977	42	1.3415	1.5408	2.5567	2.7560
16	1.0124	1.3000	2.9577	3.2543	43	1.3486	1.5458	2.5491	2.7466
17	1.0352	1.3253	2.9247	3.2148	44	1.3554	1.5506	2.5424	2.7376
18	1.0566	1.3405	2.8948	3.1787	45	1.3620	1.5552	2.5357	2.7289
19	1.0760	1.3547	2.8675	3.1456	46	1.3684	1.5596	2.5293	2.7205
20	1.0954	1.3680	2.8425	3.1151	47	1.3745	1.5638	2.5232	2.7125
21	1.1131	1.3805	2.8195	3.0869	48	1.3802	1.5678	2.5173	2.7049
22	1.1298	1.3923	2.7982	3.0607	49	1.3856	1.5716	2.5117	2.6977
23	1.1456	1.4035	2.7784	3.0362	50	1.3907	1.5752	2.5064	2.6908
24	1.1606	1.4141	2.7599	3.0133	51	1.3957	1.5787	2.5013	2.6842
25	1.1748	1.4241	2.7426	2.9919	52	1.4007	1.5822	2.4963	2.6777
26	1.1883	1.4336	2.7264	2.9718	53	1.4057	1.5856	2.4914	2.6712
27	1.2012	1.4426	2.7112	2.9528	54	1.4107	1.5890	2.4866	2.6648
28	1.2135	1.4512	2.6969	2.9348	55	1.4156	1.5923	2.4819	2.6585
29	1.2252	1.4594	2.6834	2.9177	56	1.4203	1.5955	2.4773	2.6524
30	1.2363	1.4672	2.6707	2.9016	57	1.4249	1.5987	2.4728	2.6465
					58	1.4294	1.6019	2.4684	2.6407
					59	1.4339	1.6051	2.4640	2.6350
					60	1.4384	1.6082	2.4596	2.6294

* B. J. Hart, "Significance levels for the ratio of the mean square successive difference to the variance", *Annals of Mathematical Statistics*, 13, No. 4, 1942, pp. 445-447. Reproducido con permiso de los editores.

Cuadro A-XIV

INTERVALOS DE CONFIANZA PARA LA MEDIANA*

<i>El mayor</i>			<i>El mayor</i>			<i>El mayor</i>			<i>El mayor</i>					
<i>N</i>	<i>k</i>	$\alpha \leq .05$	<i>k</i>	$\alpha \leq .01$	<i>N</i>	<i>k</i>	$\alpha \leq .05$	<i>k</i>	$\alpha \leq .01$	<i>N</i>	<i>k</i>	$\alpha \leq .05$	<i>k</i>	$\alpha \leq .01$
6	1	.031			36	12	.029	10	.004					
7	1	.016			37	13	.047	11	.008					
8	1	.008	1	.008	38	13	.034	11	.005					
9	2	.039	1	.004	39	13	.024	12	.009					
10	2	.021	1	.002	40	14	.038	12	.006					
11	2	.012	1	.001	41	14	.028	12	.004					
12	3	.039	2	.006	42	15	.044	13	.008					
13	3	.022	2	.003	43	15	.032	13	.005					
14	3	.013	2	.002	44	16	.049	14	.010					
15	4	.035	3	.007	45	16	.036	14	.007					
16	4	.021	3	.004	46	16	.026	14	.005					
17	5	.049	3	.002	47	17	.040	15	.008					
18	5	.031	4	.008	48	17	.029	15	.006					
19	5	.019	4	.004	49	18	.044	16	.009					
20	6	.041	4	.003	50	18	.033	16	.007					
21	6	.027	5	.007	51	19	.049	16	.005					
22	6	.017	5	.004	52	19	.036	17	.008					
23	7	.035	5	.003	53	19	.027	17	.005					
24	7	.023	6	.007	54	20	.040	18	.009					
25	8	.043	6	.004	55	20	.030	18	.006					
26	8	.029	7	.009	56	21	.044	18	.005					
27	8	.019	7	.006	57	21	.033	19	.008					
28	9	.036	7	.004	58	22	.048	19	.005					
29	9	.024	8	.008	59	22	.036	20	.009					
30	10	.043	8	.005	60	22	.027	20	.006					
31	10	.029	8	.003	61	23	.040	21	.010					
32	10	.020	9	.007	62	23	.030	21	.007					
33	11	.035	9	.005	63	24	.043	21	.005					
34	11	.024	10	.009	64	24	.033	22	.008					
35	12	.041	10	.006	65	25	.046	22	.006					

Si las observaciones se disponen por orden de tamaño $X_1 < X_2 < X_3 < \dots < X_N$, entonces, tenemos $100(1 - \alpha)$ por 100 de confianza de que la mediana de la población esté entre X_k y X_{N-k+1} , en donde k y α se dan anteriormente.

* Reproducido con permiso de K. R. Nair, "Table of confidence interval for the median in samples from any continuous population", Sankhyā, vol. 4 (1940). Pergamon Press, pp. 551-558.

Cuadro A-XV
DISTRIBUCIÓN DEL NÚMERO TOTAL DE R SUCESIONES EN MUESTRAS DE TAMAÑO $(n_1, n_2)^*$

$(n_1, n_2) \setminus R$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
(2,3)	.700	.500	.900	1.000															
(2,4)	.133	.100	.800	1.000															
(2,5)	.095	.333	.714	1.000															
(2,6)	.071	.286	.643	1.000															
(2,7)	.056	.250	.583	1.000															
(2,8)	.044	.222	.533	1.000															
(2,9)	.036	.200	.491	1.000															
(2,10)	.030	.182	.455	1.000															
(3,3)	.100	.300	.700	.900	1.000														
(3,4)	.057	.200	.543	.800	.971	1.000													
(3,5)	.036	.143	.429	.714	.829	1.000													
(3,6)	.024	.107	.345	.643	.833	1.000													
(3,7)	.017	.087	.283	.553	.788	1.000													
(3,8)	.012	.067	.236	.491	.745	1.000													
(3,9)	.009	.055	.200	.455	.706	1.000													
(3,10)	.007	.045	.171	.455	.706	1.000													
(4,4)	.029	.114	.371	.629	.886	.971	1.000												
(4,5)	.016	.071	.282	.500	.786	.929	.992	1.000											
(4,6)	.010	.048	.190	.405	.690	.881	.976	1.000											
(4,7)	.008	.033	.142	.333	.606	.833	.954	1.000											
(4,8)	.004	.024	.109	.279	.533	.788	.929	1.000											
(4,9)	.003	.018	.085	.236	.471	.745	.902	1.000											
(4,10)	.002	.014	.068	.203	.419	.706	.874	1.000											
(5,5)	.008	.040	.167	.357	.643	.833	.960	.992	1.000										
(5,6)	.004	.024	.110	.262	.522	.738	.911	.976	.998	1.000									
(5,7)	.003	.016	.076	.197	.424	.652	.854	.955	.992	1.000									
(5,8)	.002	.010	.054	.152	.347	.576	.793	.939	.984	1.000									
(5,9)	.001	.007	.039	.119	.297	.510	.734	.902	.972	1.000									
(5,10)	.001	.005	.029	.095	.239	.455	.678	.874	.958	1.000									
(6,6)	.002	.013	.047	.175	.392	.608	.825	.933	.987	.998	1.000								
(6,7)	.001	.008	.043	.121	.296	.500	.733	.879	.966	.992	.999	1.000							
(6,8)	.001	.005	.028	.088	.226	.413	.646	.821	.937	.984	.998	1.000							
(6,9)	.000	.003	.019	.063	.175	.343	.569	.762	.902	.972	.994	1.000							
(6,10)	.000	.002	.013	.047	.137	.288	.497	.706	.884	.958	.990	1.000							
(7,7)	.001	.004	.025	.078	.209	.383	.617	.791	.922	.975	.988	.999	1.000						
(7,8)	.000	.002	.015	.051	.149	.296	.514	.704	.867	.949	.988	.999	1.000						
(7,9)	.000	.001	.010	.035	.108	.231	.427	.622	.806	.916	.966	.994	1.000						
(7,10)	.000	.001	.006	.024	.080	.182	.355	.549	.743	.879	.957	.990	1.000						
(8,8)	.000	.001	.009	.032	.100	.214	.405	.585	.786	.900	.968	.991	.999	1.000					
(8,9)	.000	.001	.005	.020	.069	.157	.319	.500	.702	.843	.939	.980	.996	.999	1.000				
(8,10)	.000	.000	.003	.013	.048	.117	.251	.419	.621	.782	.903	.964	.990	.998	1.000				
(9,9)	.000	.000	.003	.012	.044	.109	.238	.399	.601	.782	.901	.956	.988	.997	1.000	1.000	1.000	1.000	1.000
(9,10)	.000	.000	.002	.008	.029	.077	.179	.319	.510	.681	.834	.923	.974	.992	.999	1.000	1.000	1.000	1.000
(10,10)	.000	.000	.001	.004	.019	.051	.128	.242	.414	.586	.758	.872	.946	.981	.996	.999	1.000	1.000	1.000

* C. Eisenhart y F. Swed, "Tables for testing randomness of grouping in a sequence of alternatives", Annals of Mathematical Statistics, 14(1943), p. 66. Reproducido con permiso de los editores.

Cuadro A-XV
(conclusión)
DISTRIBUCIÓN DEL NÚMERO TOTAL DE SUCESIONES

Los valores de la lista de la página anterior dan la probabilidad de que ocurrirán r o menos sucesiones. Por ejemplo, para dos muestras de tamaño 4, la probabilidad de tres o menos sucesiones es 0.114. Para las muestras de tamaño $n_1 = n_2$ mayores que 10, se puede utilizar el cuadro que se da más abajo. Las columnas encabezadas con 0.5, 1, 2.5, 5 dan valores de R de modo que ocurran R o menos sucesiones con una probabilidad menor que el porcentaje indicado. Por ejemplo, para $n_1 = n_2 = 12$ la probabilidad de 8 o menos sucesiones es de 0.05. Las columnas encabezadas con 95, 97.5, 99, 99.5 dan los valores de R para los cuales la probabilidad de R o más sucesiones es menor que 5, 2.5, 1, 0.5 por 100.

$n_1 = n_2$	0.5	1	2.5	5	95	97.5	99	99.5	Mean	Var.	s.d.
11	5	6	7	7	16	16	17	18	12	5.24	2.29
12	6	7	7	8	17	18	18	19	13	5.74	2.40
13	7	7	8	9	18	19	20	20	14	6.24	2.50
14	7	8	9	10	19	20	21	22	15	6.74	2.60
15	8	9	10	11	20	21	22	23	16	7.24	2.69
16	9	10	11	11	22	22	23	24	17	7.74	2.78
17	10	10	11	12	23	24	25	25	18	8.24	2.87
18	10	11	12	13	24	25	26	27	19	8.74	2.96
19	11	12	13	14	25	26	27	28	20	9.24	3.04
20	12	13	14	15	26	27	28	29	21	9.74	3.12
25	16	17	18	19	32	33	34	35	26	12.24	3.50
30	20	21	22	24	37	39	40	41	31	14.75	3.84
35	24	25	27	28	43	44	46	47	36	17.25	4.15
40	29	30	31	33	48	50	51	52	41	19.75	4.44
45	33	34	36	37	54	55	57	58	46	22.25	4.72
50	37	38	40	42	59	61	63	64	51	24.75	4.97
55	42	43	45	46	65	66	68	69	56	27.25	5.22
60	46	47	49	51	70	72	74	75	61	29.75	5.45
65	50	52	54	56	75	77	79	81	66	32.25	5.68
70	55	56	58	60	81	83	85	86	71	34.75	5.89
75	59	61	63	65	86	88	90	92	76	37.25	6.10
80	64	65	68	70	91	93	96	97	81	39.75	6.30
85	68	70	72	74	97	99	101	103	86	42.25	6.50
90	73	74	77	79	102	104	107	108	91	44.75	6.69
95	77	79	82	84	107	109	112	114	96	47.25	6.87
100	82	84	86	88	113	115	117	119	101	49.75	7.05

Para valores grandes de n_1 y n_2 , en particular para $n_1 = n_2$ mayor que 10, se puede utilizar una aproximación normal. La media y la varianza son

$$\frac{2n_1n_2}{n_1 + n_2} + 1 \quad \text{y} \quad \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

respectivamente. Por ejemplo, para $n_1 = n_2 = 20$, la media es 21 y la varianza es 9.74. Los percentiles 97.5 y 2.5 son $21 + 1.96 \sqrt{9.74}$ y $21 - 1.96 \sqrt{9.74}$ ó 27.1 y 14.9. La aproximación se mejora si se resta $1/2$ de los valores calculados. Los percentiles que resultan serán entonces 26.6 y 14.4. Para dos muestras cada una de tamaño n , la media es $n + 1$ y la varianza es $n(n - 1)/(2n - 1)$.

Cuadro A-XVI
NÚMEROS ALEATORIOS EQUIDISTRIBUIDOS

03 47 43 73 86	36 96 47 36 61	46 98 63 71 62	33 26 16 80 45	60 11 14 10 95
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51	24 51 79 89 73
16 76 62 27 66	50 50 26 71 07	32 90 79 78 53	13 55 38 58 59	88 97 54 14 10
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21	88 26 49 81 76
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53	23 83 01 30 30
16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 96 76
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79
33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28
18 18 07 92 46	44 17 16 58 09	79 83 86 19 62	06 76 50 03 10	55 23 64 05 05
25 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45	10 93 72 88 71
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72	93 85 79 10 75
52 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53	86 60 42 04 53
37 85 94 35 12	83 39 50 08 30	42 34 07 90 88	54 42 06 87 98	35 85 20 48 39
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04	07 74 21 19 30
50 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54	97 77 46 44 80
99 49 55 22 77	88 42 95 45 72	10 64 36 15 00	04 43 18 66 79	94 77 24 21 90
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45	99 27 72 95 14
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 60 60 44	38 68 88 11 80
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52	68 07 97 06 57
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 70	15 54 55 95 52
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45	97 60 49 04 91
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26	11 04 96 67 24
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27	40 48 73 51 92
16 90 82 60 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68	02 02 37 03 31
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86	38 45 94 30 38
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16	02 75 50 95 98
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62	48 51 84 08 32
31 96 25 91 47	96 44 33 49 13	34 86 82 53 91	00 52 43 48 85	27 55 26 89 62
66 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90	57 16 00 11 66
14 90 84 45 11	75 73 88 05 90	52 27 41 14 86	22 98 12 22 08	07 52 74 95 80
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90	49 37 38 44 59
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36	47 95 93 13 30
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 78 89 62	02 67 74 17 33
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83	52 91 05 70 74
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93	58 05 77 09 51
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16	29 56 24 29 48
26 99 61 65 53	58 37 78 80 70	42 10 50 67 42	32 17 55 85 74	94 44 67 16 94
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50	15 29 39 39 43

Fuente: Del cuadro XXXIII de Fisher y Yates: Statistical Tables for Biological, Agricultural and Medical Research, publicado por Oliver & Boyd, Ltd., Edinburgh y London, con permiso de los autores y los editores.

BORRADOR DE ARTÍCULO PARA DISCUSIÓN Y AMPLIACIÓN EN EL DEPARTAMENTO DE SISTEMAS CON VISTAS A:

1. DAR UNA INTRODUCCIÓN INFORMAL, DE SENTIDO COMÚN DE LA PROBABILIDAD Y ESTADÍSTICA.
2. AGREGANDO UNA AMPLIA Y ESCOGIDA BIBLIOGRAFÍA DE DIVULGACIÓN,
3. DE LIBROS DE TEXTO A DIFERENTES NIVELES,
4. DE ARTÍCULOS ACCESIBLES A ALUMNOS DE LICENCIATURA Y MÁS AVANZADOS.
5. DE TEXTOS Y ARTÍCULOS ESPECIALIZADOS EN DIVERSAS RAMAS DE LAS APLICACIONES EN INGENIERÍA.
6. UNA GUÍA DE LAS DIVERSAS Y MÁS COMUNES METODOLOGÍAS Y PROBLEMAS MÁS AVANZADOS DE ESTADÍSTICA (TRADUCIENDO EL ARTÍCULO ADJUNTO).
7. UNA GUÍA DE LAS DIVERSAS Y MÁS COMUNES METODOLOGÍAS Y PROBLEMAS MÁS AVANZADOS DE PROBABILIDAD.
8. UNA GUÍA DE LAS DIVERSAS Y MÁS COMUNES METODOLOGÍAS Y PROBLEMAS MÁS AVANZADOS DE PROCESOS ESTOCÁSTICOS.
9. UNA LISTA DE LOS PRINCIPALES PROBLEMAS EN CADA UNA DE LAS INGENIERÍAS RESUELTOS CON PROBABILIDAD ESTADÍSTICA Y PROCESOS ESTOCÁSTICOS.
10. UNA DESCRIPCIÓN DE LA IMPORTANCIA SOCIAL, CULTURAL, PROFESIONAL, CIENTÍFICA DE ESTOS MÉTODOS EN LA ACTUALIDAD.
11. UNA DESCRIPCIÓN DE LA HISTORIA INTELECTUAL, FILOSÓFICA Y POPULAR DE ESTOS TEMAS. Y SU IMPACTO EN LAS SOCIEDADES DEL PASADO.
12. UNA IDEA DE TEMAS QUE DENTRO DE LAS MATEMÁTICAS SE HAN PROBABILIZADO.
13. UNA INVITACIÓN Y MOTIVACIÓN A ESTUDIOS MÁS AVANZADOS.
14. REFERENCIA A MATERIAL EXISTENTE CON LOS MAESTROS, EN LA HEMEROTECA Y LA BIBLIOTECA Y/O YA SOLICITADO PARA LA UNIDAD.

INTRODUCCIÓN.

La multiplicidad de aplicaciones técnicas, conceptuales, en forma de modelos, en simulaciones con computadoras de todas las técnicas de Probabilidad y Estadística es un hecho fácil de comprobar. Su crecimiento explosivo, su novedad y en cierto sentido el estar de moda es también fácil de verificar. La computación ha hecho posible el desarrollo de muchas ideas sólo en germen previamente; las necesidades de dirección de economías y sistemas tecnológicos crecientemente complejos y precisos, han hecho una necesidad su desarrollo acelerado. En la cibernética, tal vez la disciplina que mejor defina el ambiente cultural moderno, los métodos estadísticos y probabilísticos son esenciales. Desde obreros calificados hasta presidentes de la República, el enfrentamiento a estudios e instancias donde es necesario entender, si no es que generar estudios de este tipo, son frecuentes: médicos, lingüistas, físicos, psiquiatras, ingenieros electrónicos, administradores, políticos (varios estadísticos lo han sido como Harold Wilson, ex-primer ministro de Inglaterra) premios nóbeles, economistas, investigador de operaciones, ingenieros industriales, matemáticos, corredores de bolsa, jugadores empedernidos, aseguradores, cualquier individuo con un mínimo de conciencia política, buscadores de tesoros o de petróleo y minerales, ingenieros civiles especialmente en el México sísmico, diseñadores o supervisores de plantas, todos los que quieran vender algo a nivel industrial, optimizadores de sistemas de control, cualquiera que tenga que usar termodinámica o mecánica cuántica, fisiólogos, especialmente los de nueva cuña con complejos aparatos para explorar y comprobar hipótesis, actuarios, financieros, consumidores en alerta, expertos en comunicación terrestre o extraterrestre, biólogos, químicos o químicos en la industria, especialistas en espionaje o política, contraespionaje industrial o músicos electrónicos, asociaciones de colonos, responsables de oficinas de compras o de mercadeo, líderes estudiantiles, astrofísicos, etc.

Tal vez una medida de obsolescencia y desaptación social de muchos individuos, especialmente profesionistas, sea su nivel de ignorancia de lo mucho que se ha hecho en probabilidad y estadística; por otro lado, este desarrollo multitudinario de técnicas, conceptos y usos, ahuyenta a mucha gente y con razón, por temor a perderse en un cúmulo de tecnicismos inasimilables; por eso, tal vez no se estudia o profundiza en estos tópicos enriquecedores.

Muchas de las controversias a nivel inter-industrial, profesional, político, internacional se justifican (o se tratan de justificar) a partir de "los hechos crudos" que se expresan en forma de estadísticas y modelos probabilísticos; las controversias se reducen en ocasiones a una abogacía estadística.

La divulgación de estos temas a nivel profesional, no digamos popular, es lamentablemente muy reducida. Y así se observan personas que ocupan puestos de importancia en la industria en el gobierno, en centros de investigación que cuando mucho y con suerte, han llevado un curso de estadística elemental, cuyos conocimientos han abandonado desde hace mucho y que no tienen la menor idea no digamos de los 'intestinos' de las nuevas técnicas sino, siquiera, de para que sirven.

Algunas de las metas que la sociedad se ha impuesto por mucho tiempo son las de predecir al azar, predecir 'algo' de lo impredecible del ambiente físico; las enfermedades, crisis económicas, inundaciones, pestes, escasez de alimentos, azar en la actividad económica familiar, trabajo inestable para todos, calidad y discontinuidad de servicios públicos y de los servicios y productos industriales, la calidad del ambiente, etc. 'hacer improbable lo que nos hace mal y muy probable lo que nos haga bien'. Pero esto, que se puede llamar un ~~meta~~ ^{meta}-~~meta~~ ^{meta} estadístico es inaccesible sin herramientas estadísticas, además de todas las demás.

El uso de medidas estadísticas para detectar la productividad real de secciones, industrias, ramas industriales y la economía en su conjunto y su importancia para mejorar el funcionamiento de estas es muy común y de gran importancia para superar los problemas industriales y sociales en México.

Por esto consideramos que paralelo al curso en que se estudia el máximo de probabilidad y estadística, la referencia a la lectura de material accesorio, es necesario; por cultura profesional general y como aperitivo para realizar estudios posteriores, sabiendo que este tipo de habilidades no sólo son útiles socialmente sino que hay escasez de gente preparada en estos temas.

Una de las ventajas al usuario de estas técnicas es que se presentan en forma de paquetes estadísticos disponibles en computadora, lo que los hace de fácil manejo. Sin embargo, su uso inteligente requiere de una compenetración profunda con la teoría y "ensuciarse las manos" abundantemente con ejercicios y problemas concretos. Sin embargo dadas las limitaciones de tiempo, lo que aquí se estudia no permitirá siquiera comprender de que tratan un buen número de estos programas. Lo que deseamos es que las lecturas subsanen en algo que consideramos una deficiencia, tal vez inevitable.

En México, al igual que en el mundo entero, la tradición estadística se remonta a sus primeros pobladores; se sabe que el Rey Xólotl, del grupo Chichimeca, realizó un conteo de la población de su gente en Nepohualco, (etimológicamente 'lugar en que se cuenta'). Como abundan los poblados llamados Nepohualco es de suponerse que abundaron los censos y algunos de los problemas técnicos a los que todavía nos enfrentamos. (Cf. Andrés García Pérez. Elementos de Método Estadístico. UNAM).

Las concepciones y prácticas asociadas al azar han absorbido el tiempo, la cabeza y en muchos casos el corazón de grandes personalidades históricas; también son parte integrante de la conciencia y subconciencia popular. No es, pues, de extrañarse que su estudio se vea a la vez dificultado y facilitado, por la enorme riqueza acumulada y en constante creación que esconden estos temas.

La idea del azar ha sido discutida por muchos individuos incluyendo a algunos de los más conocidos por sus aportaciones a las ciencias y a las humanidades. Muchas de sus consideraciones han sido rescatadas y puestas en forma útil por la presente Teoría de la Probabilidad y de la Estadística Matemática. Pero otras concepciones populares o tradicionales no aparecen. Algunas de las relaciones frecuentes con otros temas generales son: Azar como coincidencia de causas; como lo absolutamente fortuito, lo espontáneo lo que no tiene causa; la correspondencia del azar con la causalidad; determinismo filosófico o determinismo científico; la relación del azar con el destino, la providencia y la predestinación; el azar, la necesidad y el diseño o propósito en el origen de la estructura del mundo; causa y azar en relación al conocimiento y a la opinión; la teoría de la probabilidad, el control del azar o de la contingencia por el arte; el azar y la fortuna en la vida del individuo; azar y fortuna en política e historia.

Si alguien está interesado en ver que dijeron Cervantes, Shakespeare, Marx, Virgilio, el Antiguo y el Nuevo Testamento, etc. sobre estos temas se debe consultar la Colección:

The Great Ideas
A Syntopicon of Great Books of the Western World
Encyclopaedia Britannica Inc.
Volumen I. Bajo el inciso Chance

donde hay una recopilación de referencias a estos temas y un excelente artículo de síntesis.

Existe una corriente filosófica que afirma que el azar se puede 'domar' para alcanzar fines humanos. Tal vez esa idea subyacente en toda ciencia desinteresada, genuina. En parte es la idea de muchos reformadores o revolucionarios sociales. Pero también lo es de grupos de científicos o ingenieros que trabajaban y trabajan en la construcción de mecanismos de control, creando nuevas habilidades, aprovechando propiedades probabilísticas implícitas en los sistemas. Por ejemplo, la disciplina de control adaptivo.

El conocimiento de las condiciones probabilísticas es básico en muchos sistemas, por ejemplo, en la construcción de edificios en lugares de alta sismicidad; en reactores nucleares, en la comprensión de los fenómenos de distribu-

ción de recursos naturales, etc. para garantizar el buen servicio y evitar ca tástrofes. ,

Lo que es seguro es que esta idea, la de 'domar el azar', está directamente - relacionada con el trabajo humano, como lo atestigua la sección dedicada al mismo en la colección:

The Great Ideas
A Syntopicon of the Great Books of the Western World
Encyclopaedia Britannica, Inc.
Volumen I, Inciso Labor.

También con el concepto de madurez psicológica ejemplificado por este popu- lar dicho europeo: "Siembra acción y cosechará hábito, siembra hábito y co- sechará carácter, siembra carácter y cosechará tu suerte".

El punto de vista estadístico está profundamente imbricado en las actuales prácticas y teorías tanto en ciencias sociales, humanas como en ingeniería; Norbert Wiener, uno de los mejores matemáticos de nuestra era que contri- buyó y mejoró las aportaciones de Einstein en la Teoría Probabilística del movimiento Browniano y que trabajó en México en la creación de la Ciberné- tica, afirma en su libro, ya clásico sobre este tema, que este siglo es incom- preensible sin el uso extendido de los conceptos no determinísticos. Un exce- lente libro que colecciona artículos no matemáticos sobre Cibernética, inclu- yendo entre sus autores el gran Kolmogoroff, menciona el uso de los concep- tos de Probabilidad y Estadística en la comprensión del misterioso funciona- miento del cerebro humano. (Introducción e Historia de la Cibernética. A.V. Jramoi. Grijalbo).

Algunos libros, artículos y revistas de Divulgación que recorren muchas aplicaciones (algunas muy curiosas) son los de Émile Borel - Probabilities and Life (Dover) El muy popular libro How to lie with Statistics de Darrel Huff, ilustrado por Irving Geis (W.W. Morton Company, Penguin) presenta 'pecados' estadísticos que no por frecuentes son menos deshonestos y permi- ciosos; escrito en 'norteamericano', es muy ameno.

M.J. Momey - Facts From Figures (Pelican) (escrito por un ex-diputado in- glés, laborista y con el proverbial humor británico), por ejemplo dice que "la estadística se ha usado -de hecho se sigue usando- para permitir a los amos saber que tan lejos pueden llegar, sin causar problemas, al asaltar los bolsillos de sus conciudadanos. "También dice que la teoría es admira- ble para mucha de la práctica estadística, por descuidada y deshonestamente intelectualmente, no vale ni el papel en que está escrita.

This chancy, chancy, chancy, world, L. Rastrigin (Ed. Mir Moscú) enfatiza todo el tiempo la idea de que lo mejor de probabilidad y estadística para construir y superar herramientas y habilidades está por venir. (Excelente libro para entender estadística y probabilidad más avanzada). Ernst Wagemann,

El número detective (Fondo de Cultura Económica) cuyo autor es un chileno que llegó a tener puestos importantes en 'planeación económica' en la época previa al fascismo, en Alemania, proporciona elementos de juicio para analizar problemas macroeconómicos desde el punto de vista técnico-estadístico.

A los que tengan especial predilección por deshacerse de dinero jugando a las cartas les recomendamos el libro del Dr. Edward Thorpe. Beat the Dealer (Vintage Books) que trata sobre el "21" y su personal y ya larga historia de combate matemático-legal-estadístico-monetario con la 'mafia' que controla los casinos en E.U., el Caribe, etc.; también el artículo de 'Poder' en el número de Julio de 1978 de Scientific American.

Estos temas, que pueden considerarse frívolos o sin importancia, nos enfrentan a un problema endémico en países como el nuestro y que tiene todos los visos de crecer en el futuro: el del fraude cometido por 'genios financieros', con sus fórmulas mágicas de enriquecimiento, en incautos pequeños y medianos inversionistas, que desconocen técnicas estadísticas mínimas.

El potencial de fraudes en muchos tipos de servicios ofrecidos a simples ciudadanos en países tan comercializados es enorme. Aún en estados en que deberían saber más, como en E.U., estos delitos son comunes. (Cf. Fads and Fallacies in the name of Science - Martin Gardner. Dover. Por cierto que algunos de estos engaños hacen de las suyas, impunemente, aquí)

El libro de Marcel Boll, L'Exploitation du Hasard (Presses Universitaires de France) es accesible y muy variado, encontrando en él algunos usos esenciales de Probabilidad en la Física y Astronomía. Todos estos libros son de fácil lectura.

Finalmente de los libros que conocemos de divulgación, para no hacer muy larga la lista y así bloquear en vez de estimular su muy agradable lectura, recomendamos el Lady Luck de Warren Weaver (Double day Anchor Original) co-creador de la Teoría Estadística de la Información con Claude Shannon y que además ha ganado el premio Kalinga de la ONU por sus méritos como divulgador de la Ciencia.

Recomendamos las revistas NATURALEZA (antes FISICA) hecha en la UNAM que le ha merecido a su director Luis Estrada el mismo premio mundial y 'Ciencia y Desarrollo del CONACYT que contiene artículos científicos y

técnicos de divulgación que en ocasiones incluyen métodos o temas estadísticos. La Revista Matemática en su primera o segunda época. Y el Scientific American que tiene ya un largo historial (más de 100 años) produciendo excelentes artículos al día, introductorios a muchas ramas de las ciencias y la tecnología.

Anexamos una lista, incompleta algo arbitraria y de muy variados artículos sobre temas relacionados de algún modo (a veces en forma bastante indirecta) con probabilidad y estadística.

La revista de The American Statistician presenta artículos de interés general para la grande y creciente comunidad de especialistas, maestros y alumnos de Estadística. En el número de febrero de 1978, presenta una breve historia de los métodos gráficos en Estadística escrito por Beriger y Pohyn, desde la antigüedad hasta 1977, observándose la creciente y multiplicadora influencia de la computación.

Esperamos proporcionar posteriormente una bibliografía más completa, de mayor nivel técnico y más precisa en cuanto a aplicaciones; para ello usaremos el Engineering Index y Applied, Science and Technology Index entre otros medios de información actualizada. También, posiblemente el SECOBI del CONACYT.

Por lo pronto anexamos unas páginas de estos índices para ilustrar solamente que la incorporación de este curso en el currículum se ve más que justificado por la práctica en ingeniería vigente de estos métodos en todos los países más industrializados (de cualquier sistema político) y crecientemente en México. También para inducirlos a explorar cualquier tema (no sólo de estadística o matemáticas) de su interés para que lo desarrollen conscientemente, con los medios modernos de información, para que el resultado sea mejor.

Es claro que muchos de los artículos referidos y sintetizados en las hojas de los índices adjuntos se ven y son muy técnicos. Pero si revisan algunos de los otros temas expuestos se encontrarán que algunos artículos son accesibles y útiles inmediatamente al salir de la carrera.

El libro A Guide to Probability Theory and Application de Derman, Gleser y Olkin, sintetizó en muchos de sus problemas de tarea una gran cantidad de artículos originales (igual que el Feller). Uno de sus fuertes es el tener muchas notas históricas (al igual que el de Andrés García Pérez) y una excelente bibliografía más especializada que recomendamos sobreleer, (al igual que el muy importante libro de Feller I).

A los que quieran meterse en honduras y prepararse a entender, seguir y/o hacer investigación actual en probabilidad y procesos estocásticos puros les recomendamos el "Hoyo Negro" llamado Probability de Loeve (uno ó dos tomos según la edición) o, el Breiman y el Feller segundo tomo. Estos libros requieren algo de topología, Teoría de la Integral de Lebesgue y muchas ganas. El que quiera entender algo más tranquilamente estos tópicos Tucker "A Graduate Course in Probability" o el Asch de "Real Analysis and Probability". O el M. Frisz "Mathematical Theory of Statistics and Probability". Los alumnos interesados en tener una imagen global, totalizadora, actualizada de probabilidad y estadística, los referimos a los excelentes artículos de la Great Soviet Encyclopedia y de la Encyclopaedia Britannica sobre Probabilidad y Estadística. El esquema que brindan estos artículos puede también servir como mapa para futuras referencias cuando su uso sea necesario o para iniciar un estudio más serio.

Existen manuales o diccionarios que en forma concisa describen conceptos y técnicas como el 'Diccionario de Estadística' de Morice y Bertrand (CECSA) y el 'Experimental Statistics' (Handbook 91) United States Department of Commerce, National Bureau of Standards.

Un libro que permite ver alguno de los métodos estudiados en acción sin requerir demasiadas matemáticas es el de Noble, Engineering Applications of Undergraduate Mathematics. Asimismo los artículos del Scientific American.

También contamos con un vasto libro de aplicaciones de Estadística editado por Judith Tanur con la cooperación de renombrados estadísticos norteamericanos e ingleses: Statistics: a Guide to the Unknown. Es un libro de más de cuarenta artículos aplicados a ciencias sociales, salud pública, biología, geología, economía, computadoras, etc. y clasificados por los tipos de datos usados: muestreo, uso de datos existentes, encuestas y cuestionarios, experimentos, quasi experimentos. También están clasificados por los métodos estadísticos empleados:

- a) Estimación, pruebas de hipótesis, análisis bayesiano y análisis de datos.
- b) Tablas, gráficas y mapas.
- c) Por cientos y razones, standarización y ajuste.
- d) Series de tiempo y construcción de índices.
- e) Probabilidades y modelos.
- f) Muestreo y randomización.

- g) Correlación y regresión.
- h) Reducción y clasificación de datos.
- i) Tests y medición.
- j) Predicción.
- k) Toma de decisiones.

Hay varios artículos muy interesantes : Efecto de los límites de velocidad en carretera, precisión del concepto de justicia o injusticia social, sembradío de nubes y lluvia artificial, una relación inesperada entre el día del nacimiento de una persona y el día de su muerte (o cómo viven más algunos hombres bien conocidos).

Un problema apasionante con profundas relaciones con las concepciones sobre el origen, estructura y dinámica del universo es la distribución estadística de la masa en el universo y la posible existencia de "hoyos negros", pequeños volúmenes con densidades de masa altísimas, por ejemplo, un planeta como la Tierra en un 1 cm^3 o menos. (Veánse los artículos en Naturaleza y Ciencia y Sociedad así como el Time del 4 de Septiembre de 1978).

En México, existe una Sociedad de Control de Calidad que tiene publicaciones y reuniones especializadas. También los actuarios, carrera que se estudia en la Facultad de Ciencias de la UNAM y en la ENEP de ACATLAN, estudian y utilizan intensamente las técnicas estadísticas. Hay ya en México un grupo de profesionales especializados en el uso y/o en la investigación de las diferentes técnicas estadísticas.

LIBROS DE TEXTO

En la elaboración apresurada de estas notas tratamos de localizar el máximo de libros posibles y tratar de observar sus fortalezas y debilidades. No es descabellado pensar que algún muy buen libro se haya escabullido de nuestra vista y más aún de nuestra lista, pero es que la muestra que tomamos fue por necesidad limitada y la población madre es enorme además de creciente. Simplemente leyendo los catálogos "Books in Print" de E.U., de Alemania Occidental, de Francia, de Inglaterra, etc, que hay en la hemeroteca de la Unidad nos damos cuenta que hay un gran número de títulos muy interesantes que desconocemos. Esperamos en un futuro contar con más de estos títulos en la unidad.

Hay Tres colecciones de libros que son muy interesantes porque un mismo autor recorre los temas de probabilidad y estadística, aunque con enfoques dictados en parte por sus especialidades.

1. Hoel - Port - Stone
 - Introduction to Statistical Theory
 - Introduction to Probability Theory
 - Introduction to Stochastic Processes
2. Breiman, Leo
 - Statistics : with a view toward applications
 - Probability and Stochastic Process : with a view toward applications.
(Ambos a nivel licenciatura)
 - Probability
(A nivel de maestría - doctorado)
3. Ash
 - Probability and Statistics
 - Real Analysis, Measure and Probability
 - Information Theory
y anuncia Stochastic Processes

Un clásico en los temas de probabilidad básicamente, hecho por un gran investigador y cubriendo desde lo más sencillo hasta partes muy complejas es la colección de dos libros de Feller .- Introduction to Probability Theory. (Lo hay en Español). Su conocimiento es clave para otros temas más avanzadas. Escrito con un estilo muy vivaz y dinámico. Un libro muy usado en Estadística es el Dixon and Massey. Introduction to Statistical Analysis.

Su nivel no es muy elevado y cubre introducciones a muchos temas más avanzados de estadística. Estos provienen básicamente de experimentación agrícola, en la que los autores son expertos.

Con más énfasis en probabilidad y un poco más avanzado pero cubriendo casi el mismo número de temas que el anterior; escrito y organizado en forma muy agradable, sencilla es el texto de Miller and Freund. Su énfasis es más industrial que agrícola y sus atractivos muchos. Incluye métodos útiles para investigación de operaciones.

Un libro especialmente referido a la industria química es el de Owen L. Davies Métodos Estadísticos Aplicados a la Investigación y a la Producción. Además de ser un trabajo de equipo, surgió de experimentos estadísticos innovadores en el Imperial Chemical Industries Ltd de Inglaterra.

Engineering Statistics de Bowker - Lieberman enfatiza la concepción de la Estadística como Teoría de decisiones, e ilustra muchos de estos procesos en la industria. Escrito por un famoso estadístico y un autor de un muy popular

libro de Inv., de Operaciones. De los libros accesibles a este nivel y especializado en Físico-Química es el publicado por la editorial soviética MIR Tratamiento Matemático de datos Físico-Químicos, V.P. Spiridonov y A.A. Lopatkin.

El libro Elementos de Método Estadístico de Andrés García Pérez, publicado por la UNAM, aunque tiene casi 35 años de estar en el mercado sigue siendo útil y fácil.

Hay varios libros escritos con intereses específicos como el de Petr. Beckmann - Probability in Communication Engineering.

Como preparación para este así como para los libros de Mecánica Estadística sugerimos el libro Probabilidad y Estadística de Louis Maisel (tipo Schaum's) que tiene material de procesos estocásticos.

Léase la bibliografía comentada al final del Libro de Chou Análisis Estadístico que aunque un poco pasado de tiempo refiere a varios libros clásicos e incluye algunos temas no considerados aquí.

También insistimos en que la bibliografía a la que refiere el libro de Derman, Gleser y Olkin "A Guide to Probability Theory and Applications" contiene muchos temas de alto interés y muy diversos.

Finalmente, así como los efectos objetivos y los beneficios derivados del uso de Probabilidad y Estadística escapa, aún su enunciación, a cualquier intento de encuadramiento, los efectos subjetivos en los procesos mentales más íntimos son muy extensos. Su mayor soltura, flexibilidad los hace más atractivos y más cómodas, nos quedan mejor que otras prendas menos versátiles. Así, han surgido en los últimos 2 decenios probabilizaciones inesperadas: de los conjuntos de Cantor a los conjuntos borrosos (fuzzy sets) de Zadeh, (con lo que se obtienen topología borrosas, relaciones borrosas, sistemas borrosos, etc.), ($a \in B$ con cierta probabilidad en vez de $a \in B$ o $a \notin B$), de las pruebas lógico-matemáticas con los "imposibles" descubiertos por Gödel, a pruebas lógico-matemáticas y por ello programas que obtienen su resultado con cierta probabilidad, como Rabin y Chaitin. Ya se mencionó su impacto en la fortaleza analítica-exacta de Teoría de los números. "El pensamiento puro se probabiliza".

"Application of probability and statistics" en Summer Conference for College Teachers on Applied Mathematics. University of Missouri, 1971.

APPLICATIONS OF PROBABILITY AND STATISTICS

Lecture 13: Some Classes of Problems in Statistical Methods

Introduction

For the past two weeks you have been considering the applications of statistics, more commonly called statistical methods. I wish to consider, during this lecture, the question of what statistical methods are all about. It should be quite clear that the central activity of statistical methods is data analysis. It is undoubtedly equally clear that the underlying theory rests heavily upon the theory of probability; however, I would not wish to go so far as to say that all statistical analysis must be based upon probability models. In order to organize our thinking about the nature of statistical methods, I will proceed as follows. First, I would like to describe what I consider to be several real problems arising in research, development, and business. Secondly, I wish to label statistical problems in statistical methods as falling under one of three very broad classes. Finally, I wish to make some comments about my viewpoint on where I think statistical methods are heading.

Some "Real Life" Problems

Problem Number 1

The manager of a product line department in an electronics company has instructed his inspection department to install procedures to guarantee that the outgoing product will be of high quality. The chief inspector has put into effect an acceptance sampling plan which calls for taking a sample from each lot produced. The sample is tested and the lot is either rejected or accepted, depending upon the test results from the sample. The problem is how to devise a satisfactory sampling plan and decision rule for accepting or rejecting the lot.

Problem Number 2

A battery manufacturer has signed a contract to supply the fuel cells for a generation of space vehicles, conditional upon meeting the reliability requirements. The contract calls for the manufacturer to establish at a 95% confidence level that there is a .9999 probability that any given fuel cell will meet the engineering specifications called for by the designers of the space vehicles. The problem confronting the battery manufacturer is to devise a sampling scheme and statistical analysis which will convince his customers that he has met the reliability requirements.

Problem Number 3

A psychology researcher wishes to compare the relative effects of four different types of psychological reinforcement upon the performance of certain manual tasks. For his work he has available volunteer school children from three different elementary schools as well as four graduate students to assist him in administering the reinforcement. His problem is to design the study to compare the different types of reinforcement, adjusting as far as possible for the effects of different schools, different classes, and different workers. He also has the problem of analyzing the resulting data in a meaningful fashion.

Problem Number 4

An English professor has become interested in the possible association between English ACT scores of entering college freshmen and their resulting course score in their first course in English composition. He has obtained the English ACT scores and the final course scores for a freshman class of 723 students. He now wishes to characterize in some way the association between the two scores.

Problem Number 5

A research laboratory of a rubber company has been conducting research upon the tear strength of a particular rubber product. They have conducted an experiment in which they have controlled the percentages of components A and B and have then measured the tear strength of the resulting rubber material. Using the data obtained from the experiment, they wish to devise a method of predicting the tear strength, given the percentages of components A and B.

Problem Number 6

A farm economist is studying the trends and fluctuations in pork prices in the State of Oklahoma. He has a record of these prices for the past thirty years. He wishes to perform some sort of analysis which will provide him not only with some tenable theories about what happened in the past, but which will also enable him to predict pork prices for the immediate future.

Problem Number 7

The city planner of a small town is making an effort to foresee the future - transportation needs of the city in which he is employed. As a first step, he wishes to determine the load on the main traffic arteries of the city. He has authorized a survey to estimate the average vehicles per day on each of the

streets of interest. The natural desire is to obtain a result that is as close as possible to the true figure.

Problem Number 8

An electronics manufacturer measures and records 23 different variables or parameters on each of the electronics components produced. Recognizing the interdependence of many of these measurements, a process engineer suggests that it should be possible to combine many of the variables and parameters and to reduce the number of variables measured to a much smaller set, without the loss of any information. He has measured all 23 variables on a total of 115 components and wishes to analyze the data to see whether it is possible to reduce the number of variables.

Problem Number 9

A researcher in zoology has recorded 15 physical measurements on 100 rats each from two recognized species. Upon the basis of any one measurement, it is impossible to distinguish between the rats of the two species. He is interested in whether it is possible to devise some sort of index which takes into account all the measurements simultaneously and allows one to distinguish between the two species, purely upon the basis of the physical measurements.

Not all problems encountered in the application of statistical methods are represented by this list; however, we believe that the list does span a major portion of the fields of application. In some senses the problems seem to be quite different and the appropriate methods seem to be unrelated. In the next section, however, we will suggest that all of these problems fit within one or more of three general classes of problems.

Some Important Classes of Problems in Statistical Methods

If one were to classify the problems discussed in the previous section by titles commonly encountered in the tables of contents or indexes of most books in statistical methods, we suggest that the nine problems would probably be classified by the following nine titles :

1. Acceptance Sampling
2. Reliability Analysis
3. Design of Experiments
4. Correlation Analysis
5. Multiple Regression
6. Time Series Analysis

7. Survey Sampling
8. Factor Analysis
9. Discriminant Analysis

These labels are useful and do serve to delineate certain areas of work in statistics; however, as can be readily perceived, this is only a partial list. A complete list of the problem areas in statistical methods is too long to permit comprehension of what statistical methods is all about. We would like to suggest that statistical methods can be classified as belonging to one or more of the following three classes of problems :

1. Fitting of models to data
2. Goodness of fit of completely specified models
3. Goodness of fit of partially specified models

We believe that these three subjects provide a useful framework by which one can think intelligently about problems in statistics and solutions to such problems. Clearly, the central feature which characterizes modern statistics is the use of a mathematical or probabilistic model as the basis for the analysis of data. It seems almost axiomatic, then, that the central problems are the fitting of such models to the data obtained and the judgment of how well the models fit the data. In judging the goodness of fit of models to data, it has seemed useful to us to distinguish between completely specified models and partially specified models. In most serious statistical applications we use only partially specified models, although completely specified models are usually used in introducing basic statistical techniques.

Let us now consider the problems presented in the previous section and consider how they may be viewed as falling under one or more of these general classes of problems :

Problem Number 1

The model likely to be assumed for a sample drawn from a production process is that of independent Bernoulli trials with constant probability p . The occurrence of a defective item correspond to a Bernoulli trial outcome of zero. The problem of estimating the lot fraction defective is then viewed as a problem of estimating the p of the Bernoulli trials, given the acceptance sampling results. The model which we are using is only partially specified, since the p is regarded as unknown. The problem of deciding whether to accept or reject the lot can be viewed as one of comparing the goodness of fit of the sample data with a class of models with p large as opposed to a class of models with p small. If the sample data are better fitted by a partially specified model with unknown p , but smaller than some given value, then we will be led to accept the lot. If, on the other hand, the sample data are better fitted by a partially specified model with large p , then we will be led to reject the lot.

Problem Number 2

A model for the reliability problem might be that the items sampled by the manufacturer constitute a random sample from some mathematical distribution, the degree of specification depending upon whether or not the form of the distribution is assumed. Generally, even if the form of the distribution is assumed, the model is still only partially specified because the parameter values of the distribution will not be assumed to be known. In order for the reliability analysis to have real weight, and to be truly convincing, there must be some indication that the model is an appropriate one. We are therefore led again to the problem of fitting the unknown parameters of the model to the sample data. The question of whether or not the process reliability meets the contractual requirements can be viewed as a matter of the goodness of fit of the data to a partially specified model. We are interested in what fraction of the population exceeds a given engineering requirement. We therefore set out to determine the goodness of fit of the data by a mathematical distribution, only partially specified, in which the area above the engineering specification exceeds the required reliability.

Problem Number 3

In discussing the problem of comparing four different reinforcement conditions, I am going to assume that all of you are familiar with the analysis of variance and are familiar with the use of an F test, given by the treatment mean square over the error mean square, to test the hypothesis that the treatment means are all equal. This sort of test is usually justified by the use of a model which involves and overall effect, a treatment effect, a block effect plus other identifiable effects, plus an error term which is normally and independently distributed with mean zero and constant variance σ^2 . I would like to suggest that the F test for no treatment differences can be viewed simply as judging how well the data are fitted by a model in which there are no treatment terms. The model is a partially specified model, of course, because the values for the overall effect, the block effects, and the error variance are not specified.

Problem Number 4

In looking at the ACT scores and the final course scores, the model for the data might be that the pairs of scores constitute a random sample from a bivariate normal distribution with unspecified means, variances, and correlation coefficient. The model is then fitted to the data by the process of what is commonly called estimation of parameters, and judgments about the degree of association between the two measured scores may be viewed as comparing the goodness of fit of the data to a bivariate normal with zero correlation with a bivariate normal with a nonzero correlation.

Problem Number 5

In trying to get an equation to predict the tear strength from input variables, it is very clear that this problem fits in the first class of problems described, namely, that of fitting models to data. It should also be clear that one of the major problems in regression analysis is the forming of opinions about how well the equations fit the data. Since the terms in the regression equations, as well as the value of the error of variance, are generally unspecified, the model is again an example of a partially specified model.

Problem Number 6

The degree of specification for time series analysis varies widely from a model in which we assume only something like variance stationarity or covariance stationarity to a model in which the basic structure (apart from unspecified constants) for the data-generating mechanism is assumed. In the loosely specified case the fitting of the model to the data usually amounts only to the calculation of some simple statistic, such as lag correlation coefficients, or a spectrogram. In the more completely specified models the fitting of the model to the data will involve estimation of the unspecified parameters in the data-generating model. In all cases we must form some opinion about the goodness of fit of the data to the model assumed.

Problem Number 7

In the case of a sample from an existing population, the model is very loosely specified. It usually consists only of the assumption that the individuals in the sample constitute a random sample of a fixed finite population. In this case fitting of the model to the data usually consists only of the calculation of simple things such as the mean or the variance. The goodness of fit may involve problems like judging whether or not the population total exceeds a given value. For example, in the traffic survey referred to, one might be interested in whether or not the data is well-fitted by a model with a traffic load greater than a specified value.

Problem Number 8

If a factor analysis is done without consideration of some underlying model, it remains purely and simply descriptive statistics and only falls marginally within the limits of this lecture. If, however, there is an underlying probability model, then the problems of factor analysis are again those of fitting the sample data to the model and judging the goodness of fit to the model.

Problem Number 9

The model for discriminant analysis may be that the sample data have come from two different mathematical distributions with unspecified parameters. Upon the basis of the data, we wish to fit these distributions to the data, and the resulting problem of discrimination is then a matter of which distribution best fits which observations.

Perspective

To attempt to describe the whole of statistical methods as falling within three classes of problems, is in one sense, foolhardy. Obviously, the description of all statistical activity is an immense task and is, in itself, a statistical problem. However, we believe that the framework we have presented does provide a good perspective for comprehending the whole of the application of statistical methods. We believe that it is consistent with the ideas of the scientific method. We believe that it is useful in helping people to conceive of statistical methods are viewed in this fashion, rather than as all-purpose remedies prescribed by your local statistician, it becomes apparent that it is very difficult to do a good and effective job in the application of statistical methods.

What about people with insufficient training who have this broad perspective? What about people who wish to use some prescribed statistical methods for the set of data which they have in hand? Should they be encouraged to apply statistical methods in such a cookbook fashion? There are those who would say no. However, it is my own opinion that the routine application of statistical methods, many times by people with only marginal understanding of what they are about, has proven to be useful for scientific investigation. My answer is, therefore, that they should be encouraged to apply statistical methods.

However, as research workers become better trained, they will no longer be satisfied with routine approaches to the analysis of data. It is my personal belief that they will be better satisfied in looking at statistical methods as I have described them here today, even though the resulting analysis may seem to have something more of an ad hoc nature.

I believe that in future years we will see far more statistical analysis than we have witnessed to date. I believe that the users of statistics will become more and more sophisticated and that the general philosophy and attitude toward the application of statistical methods will necessarily change and become more sophisticated.

LEVIN, Jack, "Aplicación de métodos estadísticos a problemas de investigación" en Estadística para las ciencias sociales. Edit. Limusa, 1978.

La Parte III del texto contiene varias técnicas estadísticas que se pueden aplicar a los diferentes problemas de la investigación social. Los Capítulos 8, 9 y 10 presentaron las diversas técnicas utilizadas para determinar si las diferencias muestrales obtenidas son estadísticamente significativas o sólo un simple producto del error de muestreo. Las técnicas del Capítulo 11 tienen por objeto determinar el grado de asociación, la correlación entre dos variables.

Como se ha hecho notar, a través de todo el texto, cada técnica estadística tiene un conjunto de hipótesis para su correcta aplicación. En la selección de las técnicas, cualquier investigador deberá tener en cuenta varios factores, tales como:

1. si el investigador busca contrastar diferencias estadísticamente significativas, el grado de asociación, o ambos;
2. si el investigador ha alcanzado el nivel de medición nominal, ordinal o por intervalos de las variables en estudio;
3. si las variables que se están estudiando están o no distribuidas normalmente en la población de donde fueron extraídas; y
4. si el investigador está estudiando muestras independientes o la misma muestra medida más de una vez.

El presente capítulo proporciona una serie de situaciones hipotéticas de investigación en las que se especifican los criterios anteriores. Se pide al estudiante que escoja la técnica estadística más apropiada para cada situación de investigación de entre las siguientes pruebas que se vieron en la Parte III del texto:

1. la razón t
2. el análisis de varianza

3. la χ^2 cuadrada
4. la prueba de la mediana
5. el análisis de varianza en una dirección de Kruskal-Wallis
6. el análisis de varianza en dos direcciones de Friedman
7. la r de Pearson
8. el orden de los rangos de Spearman
9. gamma de Goodman y Kruskal
10. phi
11. el coeficiente de contingencia
12. la V de Cramér

La Tabla 12.1 (p. 244) sitúa cada técnica estadística con respecto a algunas de las suposiciones importantes que se deben tener en cuenta para su correcta aplicación. Mirando las columnas de la tabla nos encontramos frente a la primera decisión importante relacionada con la selección de una técnica estadística: ¿Deseamos determinar si existe o no una relación? Las pruebas de significancia estudiadas en los Capítulos 8, 9 y 10 tienen por objeto determinar si una diferencia muestral obtenida refleja una diferencia poblacional verdadera. O acaso ¿buscamos establecer la fuerza de la relación entre dos variables? Esta es una cuestión de correlación a la que nos podemos dirigir por medio de las técnicas estadísticas presentadas en el Capítulo 11. Los subtítulos de las columnas de la Tabla 12.1 indican que un investigador que decide emplear una prueba de significancia en lugar de una técnica de correlación debe saber si está estudiando muestras independientes o la misma muestra medida más de una vez.

Los renglones de la Tabla 12.1 dirigen nuestra atención hacia el nivel al que están medidas nuestras variables. Si hemos logrado el nivel de medición por intervalos bien podríamos pensar en el empleo de una técnica paramétrica como t , F , etc. Sin embargo, ya sea que hayamos llegado al nivel de medición nominal o al ordinal la elección se limitará a varias alternativas no paramétricas.

Al final del capítulo se pueden encontrar las soluciones a las siguientes situaciones de investigación.

SITUACIONES DE INVESTIGACION

Situación de investigación 1

Un investigador realizó un experimento para determinar el efecto de la edad de un conferencista sobre la preferencia de los estudiantes para escuchar sus conferencias. En una situación normal, dentro del salón de clases, se dijo a 20 estudiantes que la administración quería saber acerca de sus preferencias respecto a una próxima serie de conferencistas visitantes. Específicamente, se les pidió evaluar a un profesor que "podría venir de visita a la universidad". A todos los estudiantes se les describió del mismo modo el profesor excepto porque: a la mitad de los alumnos se le dijo que el profesor tenía 65 años de edad; a la otra mitad se le dijo que el profesor tenía sólo 25. Se pidió entonces a todos los estudiantes que indicaran su disposición

para asistir a la conferencia de dicho profesor (los datos más altos indican una mayor disposición). Se obtuvieron los siguientes resultados:

X_1 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 25 años)	X_2 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 65 años)
65	78
38	42
52	77
71	50
69	65
72	70
55	55
78	51
56	33
80	59

¿Qué procedimiento estadístico se podría aplicar para determinar si existe una diferencia significativa entre estos grupos de estudiantes con respecto a su disposición para asistir a la conferencia?

Situación de investigación 2

Un investigador llevó a cabo un experimento para determinar el efecto de la edad de un conferencista sobre la preferencia de los estudiantes para escuchar sus conferencias. En una situación normal dentro del salón de clase, se dijo a 30 estudiantes que la administración deseaba conocer sus preferencias en relación con una futura serie de conferencistas visitantes. Concretamente se les pidió que evaluaran a un profesor que “podría venir de visita a la universidad”. El profesor fue descrito a todos los estudiantes de la misma manera, sólo que a un tercio de los alumnos se les dijo que el profesor tenía 75 años de edad; a un tercio se le dijo que tenía 50; y a un tercio se le dijo que tenía sólo 25. Luego se pidió a todos los estudiantes que indicaran su disposición para asistir a la conferencia del profesor. Se obtuvieron los siguientes resultados:

X_1 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 25 años)	X_2 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 50 años)	X_3 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 75 años)
65	63	67
38	42	42

TABLA 12.1 Criterios
para escoger una
técnica estadística
apropiada

Nivel de medición	Pruebas de significancia (Capítulos 8, 9, 10)		Correlación (Capítulo 11)
	Muestras independientes	La misma muestra medida dos veces	
<i>Nominal</i>	Chi cuadrada (prueba no paramétrica para com- parar dos o más muestras)		Coficiente phi (tabla 2 X 2 no paramétrico) Contingencia y Y de Cra- mér (para tablas mayo- res de 2 X 2 no paramé- tricas) -
<i>Ordinal</i>	Prueba de la mediana no paramétrica para com- parar dos muestras) Análisis de varianza en una dirección de Kruskal- Wallis (no paramétrico para comparar tres o más muestras)	Análisis de varianza en dos direcciones de Friedman (no paramétrico para comparar la misma muestra medida por lo menos dos veces)	Orden de rango de Spear- man (no paramétrico) Gamma de Goodman y Kruskal (no paramétrica para tratar un gran número de rangos empa- tados)
<i>Intervalo</i>	Razón <i>t</i> (paramétrica para comparar dos muestras) Análisis de varianza (para- métrico para comparar tres o más muestras)	Razón <i>t</i> (paramétrica para comparar la misma mues- tra medida dos veces)	<i>r</i> de Pearson (paramétrica)

X_1 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 25 años)	X_2 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 50 años)	X_3 (Puntajes de estudiantes a quienes se dijo que el profesor tenía 75 años)
52	60	77
71	55	32
69	43	52
72	36	34
55	69	45
78	57	38
56	67	39
80	79	46

¿Qué procedimiento estadístico se podría aplicar para determinar si existe una diferencia significativa entre estos grupos de estudiantes con respecto a su disposición para asistir a la conferencia?

Situación de investigación 3

Para investigar la relación entre la ortografía y la habilidad para la lectura, un investigador aplicó exámenes de ortografía y de lectura a un grupo de 20 estudiantes seleccionados aleatoriamente de una gran población de estudiantes no graduados. Se obtuvieron los siguientes resultados (los puntajes más altos indican una mayor habilidad):

	X (Puntaje de ortografía)	Y (Puntaje de lectura)
A	52	56
B	90	81
C	63	75
D	81	72
E	93	50
F	51	45
G	48	39
H	99	87
I	85	59
J	57	56
K	60	69
L	77	78
M	96	69
N	62	57
O	28	35
P	43	47
Q	88	73
R	72	76
S	75	63
T	69	79

¿Qué procedimiento estadístico se podría aplicar para determinar el grado de asociación entre la ortografía y la habilidad para la lectura?

Situación de investigación 4

Para averiguar la validez de un determinado examen de lectura, los investigadores lo aplicaron a una muestra de 20 estudiantes cuya habilidad para leer había sido previamente colocada por rangos por su profesor. El puntaje del examen y el rango que el profesor dio para cada estudiante se enumeran a continuación:

<i>Estudiante</i>	<i>X (Puntaje de lectura)</i>	<i>Y (Rango del profesor)</i>
A	28	18
B	50	17
C	92	1
D	85	6
E	76	5
F	69	10
G	42	11
H	53	12
I	80	3
J	91	2
K	73	4
L	74	9
M	14	20
N	29	19
O	86	7
P	73	8
Q	39	16
R	80	13
S	91	15
T	72	14

¿Qué procedimiento estadístico se podría aplicar para determinar el grado de asociación entre los puntajes de lectura y la categorización del profesor?

Situación de investigación 5

Para estudiar las diferencias regionales relacionadas con el espíritu servicial hacia los desconocidos, un investigador dejó caer 400 llaves (todas las cuales habían sido marcadas y señaladas con una dirección de remitente) en los alrededores de los buzones de las regiones norte, sur, este y oeste de una ciudad. El número de llaves devueltas por región (como un indicador del espíritu servicial) se indica a continuación:

	<i>Región</i>			
	<i>Norte</i> <i>f</i>	<i>Sur</i> <i>f</i>	<i>Este</i> <i>f</i>	<i>Oeste</i> <i>f</i>
Devueltas	55	69	82	61
No devueltas	45	31	18	39
	100	100	100	100

¿Qué procedimiento estadístico se podría aplicar para determinar si estas diferencias regionales son estadísticamente significativas?

Situación de investigación 6

Para examinar la relación entre el autoritarismo y los prejuicios, un investigador administró medidas de autoritarismo (la escala F) y prejuicio (una lista de confrontación de los adjetivos negativos generalmente asignados a los norteamericanos negros) a una muestra nacional de 950 norteamericanos adultos. Se obtuvieron los siguientes resultados: de 500 entrevistados autoritarios, 350 estaban “prejuiciados” y 150 eran “tolerantes”. De 450 entrevistados no autoritarios, 125 estaban “prejuiciados” y 325 eran “tolerantes”.

¿Qué procedimiento estadístico se podría aplicar para estudiar el grado de asociación entre el autoritarismo y el prejuicio?

Situación de investigación 7

Para investigar la relación entre el año escolar y el promedio de calificaciones, los investigadores examinaron los antecedentes académicos de 186 estudiantes universitarios seleccionados sobre una base aleatoria de la población no graduada de cierta universidad. Los investigadores obtuvieron los siguientes resultados:

	<i>Año escolar</i>			
	<i>1o.</i> <i>f</i>	<i>2o.</i> <i>f</i>	<i>3o.</i> <i>f</i>	<i>4o.</i> <i>f</i>
<i>Promedio de calificaciones</i>				
MB	6	5	7	10
B	10	16	19	18
S	23	20	15	7
NA	15	7	6	2
	54	48	47	37

(*N* = 186)

¿Qué procedimiento estadístico se podría aplicar para determinar el grado de asociación entre el promedio de calificaciones y el año escolar de los alumnos?

Situación de investigación 8

Para investigar la influencia de la frustración sobre los prejuicios, se pidió a 10 sujetos que asignaran adjetivos negativos como perezoso, sucio e inmoral, para describir a los miembros de un grupo minoritario (una medida de prejuicio). Todos los sujetos describieron al grupo minoritario tanto antes como después de que habían tomado una serie de exámenes largos y difíciles (la situación frustrante). Se obtuvieron los siguientes resultados (los puntajes más altos representan un mayor prejuicio):

<i>Sujeto</i>	X_1	X_2
	<i>(Puntajes de prejuicio antes de tomar los exámenes frustrantes)</i>	<i>(Puntajes de prejuicio después de tomar los exámenes frustrantes)</i>
A	22	26
B	39	45
C	25	24
D	40	43
E	36	36
F	27	29
G	44	47
H	31	30
I	52	52
J	48	59

¿Qué procedimiento estadístico se podría aplicar para determinar si existe una diferencia estadísticamente significativa en los prejuicios antes y después de la administración de los exámenes frustrantes?

Situación de investigación 9

Para investigar la relación entre el estatus ocupacional real de un entrevistado y su clase social subjetiva (o sea, su propia identificación de clase social), se pidió a 677 individuos que indicaran su ocupación y la clase social a la que pertenecían. De 190 entrevistados con ocupaciones de estatus superior (profesional-técnico-gerencial), 56 se identificaron como miembros de la clase alta, 122 de la clase media, y 12 de la clase baja; de 221 entrevistados con ocupaciones de estatus medio (vendedores-oficinistas-trabajadores calificados), 42 se identificaron como miembros de la clase alta, 163 de la clase media, y 16 de la clase baja; de 266 entrevistados con ocupaciones de estatus bajo (trabajadores de mano de obra semi calificada y no calificada), 15 se identificaron como miembros de la clase alta, 202 de la clase media y 49 de la clase baja.

¿Qué procedimiento estadístico se podría aplicar para determinar el grado de asociación entre el estatus ocupacional y la clase social subjetiva?

Situación de investigación 10

Para investigar la influencia de la especialización universitaria en el sueldo inicial de los graduados universitarios, los investigadores entrevistaron a un grupo de estudiantes recién graduados, especializados en ingeniería, ciencias sociales o administración de empresas, en relación con sus primeros empleos. Los resultados obtenidos para estos 21 entrevistados son los siguientes:

Salarios iniciales		
<i>Ingeniería</i>	<i>Ciencias sociales</i>	<i>Administración de empresas</i>
\$ 10 500	\$ 7 000	\$ 7 500
12 300	9 500	9 000
14 000	10 000	8 000
9 500	11 000	9 300
9 000	8 500	10 500
8 500	7 500	10 000
7 500	7 000	7 000

¿Qué procedimiento estadístico se podría aplicar para determinar si existe una diferencia significativa entre estos grupos de entrevistados con respecto a sus salarios iniciales?

Situación de investigación 11

Para investigar la influencia de la especialización universitaria en el salario inicial de los graduados universitarios, los investigadores entrevistaron a un grupo de estudiantes recién graduados, especializados en ciencias sociales o en administración, en relación con sus primeros empleos. Los resultados obtenidos para estos 16 entrevistados son los siguientes:

Salarios iniciales	
<i>Ciencias sociales</i>	<i>Administración</i>
\$ 7 000	\$ 7 500
9 500	9 000
10 000	8 000
11 000	9 300
8 500	10 500
7 500	10 000
7 000	7 000
	8 000
	9 300

¿Qué procedimiento estadístico se podría aplicar para determinar si existe una diferencia significativa entre los especialistas en ciencias sociales y los especialistas en administración con respecto a sus salarios iniciales?

Situación de investigación 12

Un investigador llevó a cabo un experimento para determinar el efecto de la edad de un conferencista sobre la disposición estudiantil para escuchar sus conferencias. En una situación normal, dentro del salón de clases, se dijo a 130 estudiantes que la administración deseaba conocer sus preferencias respecto a una próxima serie de conferencistas visitantes. Específicamente, se les pidió evaluar a un profesor que “podría venir de visita a la universidad”. El profesor fue descrito igualmente para todos, a no ser porque: a la mitad de los estudiantes se le dijo que el profesor tenía 65 años de edad y a la otra mitad se le dijo que el profesor tenía sólo 25. Más tarde se pidió a todos los estudiantes que indicaran su disposición para asistir a la conferencia del profesor y se obtuvieron los siguientes resultados: de los estudiantes a quienes se dijo que el profesor tenía 65 años, 22 manifestaron su disposición para asistir a las conferencias y 43 expresaron su renuencia; de los estudiantes a quienes se dijo que el profesor tenía 25 años, 38 manifestaron su disposición de asistir a las conferencias y 27 expresaron su renuencia.

¿Qué procedimiento estadístico se podría aplicar para determinar si existe una diferencia significativa entre estos grupos de estudiantes con respecto a su disposición para asistir a la conferencia del profesor?

SOLUCION A LAS INVESTIGACIONES

Solución a la situación de investigación 1

(Razón t o prueba de la mediana)

La situación de investigación 1 representa una comparación entre los puntajes de dos muestras independientes de estudiantes. La razón t (Capítulo 8) se emplea con el fin de hacer comparaciones entre dos medias cuando se han obtenido datos por intervalos. La prueba de la mediana (Capítulo 10) es una alternativa no paramétrica que se puede aplicar cuando sospechemos que los puntajes no están distribuidos normalmente en la población o que no se ha logrado el nivel de medición por intervalos.

Solución a la situación de investigación 2

(Análisis de varianza o análisis de varianza en una dirección de Kruskal-Wallis)

La situación de investigación 2 representa una comparación de los puntajes de tres muestras independientes de estudiantes. La razón F (análisis de varianza, Capítulo 9) se emplea para hacer comparaciones entre tres o más medias independientes cuando se han obtenido datos por intervalos. El análisis de varianza en una dirección de Kruskal-Wallis (Capítulo 10) puede aplicarse como una alternativa no paramétrica cuando tenemos motivos para sospechar que los puntajes no están distribuidos

normalmente en la población o cuando no se ha alcanzado el nivel de medición por intervalos.

Solución a la situación de investigación 3
(*La r de Pearson*)

La situación de investigación 3 es un problema de correlación puesto que pide el grado de asociación entre X (habilidad en ortografía) y Y (habilidad para la lectura). La r de Pearson (Capítulo 11) puede emplearse para detectar una correlación lineal entre las variables X y Y cuando ambas han sido medidas al nivel por intervalos. Si X (habilidad en ortografía) y Y (habilidad en lectura) no están distribuidas normalmente en la población, habrá que pensar en la aplicación de una alternativa no paramétrica tal como el coeficiente de correlación por rangos ordenados de Spearman (Capítulo 11).

Solución a la situación de investigación 4
(*Rangos ordenados de Spearman*)

La situación de investigación 4 es un problema de correlación que pregunta por el grado de asociación entre X (puntajes de lectura) y Y (evaluación del profesor respecto a la habilidad para la lectura). El coeficiente de correlación por rangos ordenados de Spearman (Capítulo 11) puede emplearse para detectar una relación lineal entre las variables X y Y , cuando ambas variables han sido ordenadas o colocadas por rangos. La r de Pearson no se puede emplear pues requiere el nivel de medición por intervalos para X y Y . En el presente caso, los puntajes de lectura (X) deben ser colocados por rangos 1 a 20 antes de aplicar el coeficiente por rangos ordenados.

Solución a la situación de investigación 5
(*Chi cuadrada*)

La situación de investigación 5 representa una comparación entre las frecuencias (llaves devueltas contra llaves no devueltas) encontradas en cuatro grupos (norte, sur, este y oeste). La prueba de significancia chi cuadrada (Capítulo 10) se utiliza para hacer comparaciones entre dos o más muestras. Sólo se requieren los datos nominales. Los presentes resultados se pueden colocar en forma de tabla 2×4 , representando 2 renglones y 4 columnas. Nótese que el grado de asociación entre la tasa de devolución (X) y la región (Y) se puede medir con el coeficiente de contingencia (C) o la V de Cramér (Capítulo 11).

Solución a la situación de investigación 6
(*Coeficiente ϕ*)

La situación de investigación 6 es un problema de correlación que pregunta por el grado de asociación entre X (autoritarismo) y Y (prejuicio). El coeficiente ϕ (Capí-

tulo 11) es una medida de asociación que puede emplearse cuando los datos de frecuencia o nominales se pueden colocar en forma de tabla 2×2 (2 renglones y 2 columnas). En el presente problema, dicha tabla tomaría la forma siguiente:

Nivel de prejuicio	Nivel de autoritarismo		
	Autoritario	No autoritario	
Prejuiciado	350	120	$N = 950$
Tolerante	150	325	

Solución a la situación de investigación 7
(Gamma de Goodman y Kruskal)

La situación de investigación 7 es un problema de correlación que pregunta por el grado de asociación entre X (promedio de calificaciones) y Y (año escolar). El coeficiente gamma de Goodman y Kruskal (Capítulo 11) se emplea para detectar una relación lineal entre X y Y cuando ambas variables se han colocado por rangos y ha ocurrido un gran número de empates. En el presente problema, el promedio de calificaciones se ha colocado por rangos desde MB hasta NA y el año escolar se ha colocado por rangos de 1o. a 4o. Ambas medidas ordinales crudas han generado numerosos rangos empatados (por ejemplo, 54 estudiantes estaban en su primer año escolar; 48 el segundo, y así sucesivamente). El coeficiente de contingencia (C) o la V de Cramér (Capítulo 11) representa una alternativa en relación con gamma, la cual supone únicamente datos de nivel nominal.

Solución a la situación de investigación 8
(Razón t o análisis de varianza en dos direcciones por rangos)

La situación de investigación 8 representa una comparación antes-después de una sola muestra medida en dos puntos diferentes en el tiempo. La razón t (Capítulo 8) puede emplearse para comparar dos medias de una sola muestra ordenada en un diseño de panel antes-después. El análisis de varianza en dos direcciones de Friedman (Capítulo 10) es una alternativa no paramétrica que se puede aplicar a la situación antes-después cuando tenemos motivos para sospechar que los puntajes no están distribuidos normalmente en la población o cuando no hemos alcanzado el nivel de medición por intervalos.

Solución a la situación de investigación 9
(Gamma de Goodman y Kruskal)

La situación de investigación 9 es un problema de correlación que pregunta por el grado de asociación entre X (estatus ocupacional) y Y (clase social subjetiva).

El coeficiente gamma (Capítulo 11) es detectar una relación lineal entre X y por rangos y ha ocurrido un gran número de estatus ocupacional y la clase social subjetiva se han ordenado de “alta” a “media” y a “baja”, generando un número muy grande de rangos empatados (por ejemplo, 221 entrevistados tenían ocupaciones de estatus medio). Para obtener el coeficiente gamma, se deben reordenar los datos en forma de tabla de frecuencia como sigue:

<i>Clase social subjetiva (Y)</i>	<i>Estatus ocupacional (X)</i>		
	<i>Alto f</i>	<i>Medio f</i>	<i>Bajo f</i>
Alta	56	42	15
Media	122	163	202
Baja	12	16	49
	190	221	266

El coeficiente de contingencia (C) y la V de Cramér son alternativas para gamma que suponen sólo datos nominales.

Solución a la situación de investigación 10
*(Análisis de varianza o análisis de varianza
 en una dirección de Kruskal-Wallis)*

La situación de investigación 10 representa una comparación de los puntajes de tres muestras independientes de entrevistados. La razón F (Capítulo 9) se utiliza para

dos muestras independientes de entrevistados. La razón t (Capítulo 8) se emplea para comparar dos medias cuando se han obtenido datos por intervalos. La prueba de la mediana (Capítulo 10) es una alternativa no paramétrica que puede aplicarse cuando no podemos suponer que los puntajes están distribuidos normalmente en la población o cuando no se ha alcanzado el nivel de medición por intervalos.

Solución a la situación de investigación 12
(Chi cuadrada)

La situación de investigación 12 representa una comparación de las frecuencias (disposición contra renuencia) en dos grupos de estudiantes (aquéllos a quienes se dijo que el profesor tenía 65 años contra aquéllos a quienes se dijo que tenía 25). La prueba de significancia chi cuadrada (Capítulo 10) se usa para hacer comparaciones entre dos o más muestras cuando se han obtenido datos nominales o de frecuencia. Los presentes resultados pueden colocarse en forma de la siguiente tabla 2×2 , que representen 2 renglones y 2 columnas:

<i>Disposición para asistir</i>	<i>Condición experimental</i>		
	<i>Estudiantes a quienes se dijo que el profesor tenía 65 años f</i>	<i>Estudiantes a quienes se dijo que el profesor tenía 25 años f</i>	
<i>Dispuesto</i>	22	38	<i>N = 130</i>
<i>Renuente</i>	43	27	

Guía SAI de
probabilidad
y estadística

Tronco general

Se terminó	La edición estuvo
de imprimir	a cargo
en el mes de junio	de la Sección
del año 2002	de Producción
en los talleres	y Distribución Editoriales
de la sección	
de impresión	Se imprimieron
y reproducción de la	250 ejemplares
Universidad Autónoma Metropolitana	más sobrantes
Unidad Azcapotzalco	para reposición.



Formato de Papeleta de Vencimiento

*El usuario se obliga a devolver este libro en la fecha
señalada en el sello mas reciente*

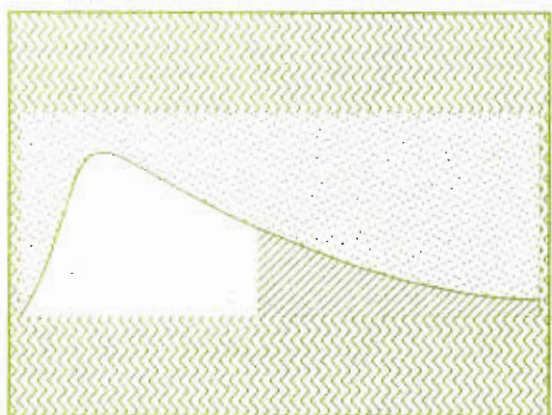
Código de barras.

2893128

FECHA DE DEVOLUCION

[illegible]

- Ordenar las fechas de vencimiento de manera vertical.
- Cancelar con el sello de "DEVUELTO" la fecha de vencimiento a la entrega del libro



0092101 05251



13.00 - \$ 13.00